DOCUMENT RESUME

ED 423 483 CG 028 708

AUTHOR Owen, K.
TITLE The Role of Psychological Tests in Education in South Africa: Issues, Controversies and Benefits.
INSTITUTION Human Sciences Research Council, Pretoria (South Africa).
ISBN ISBN-0-7969-1881-3
PUB DATE 1998-00-00
NOTE 129p.
AVAILABLE FROM Human Sciences Research Council, 134 Pretorius St., Pretoria, South Africa 0002.
PUB TYPE Books (010) -- Opinion Papers (120)
EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Ability Identification; Blacks; *Culture Fair Tests; *Educational Policy; Educational Testing; Elementary Secondary Education; Foreign Countries; Intelligence Differences; Personality Assessment; *Psychological Testing; Schools; Test Bias; *Test Use; *Values
IDENTIFIERS *South Africa

ABSTRACT

This volume examines historic, cross-cultural, and psychometric issues with regard to the use of psychological testing in South Africa. After an introduction in Chapter 1, the following chapters are: "Measurement and Evaluation in Psychology and Education"; "History of the Development of Psychological Tests," which includes intelligence, aptitude, and personality tests; "Approaches to the Assessment of Cognitive Development," which reviews the psychometric, Piagetian, and Soviet approaches, neuropsychologically based instruments, and dynamic assessment. "Psychological Testing: Criticisms, Issues and Controversies," which explores both criticisms and test bias; "Culture and Testing," which discusses the influence of culture on test performance, offers six possible solutions to selection issues, and discusses a Eurocentric versus an Afrocentric approach to testing; "The Role of Psychological Tests in South African Schools," which includes cognitive, individual intelligence, group intelligence, aptitude and proficiency, and personality tests; and "Psychological Testing in South Africa: End of the Road or a New Beginning?" (Contains 95 references.) (EMK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 483

THE ROLE OF *psychological tests in education* IN SOUTH AFRICA

ISSUES, CONTROVERSIES AND BENEFITS

K OWEN

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BH Stumpf

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

CG028708

THE ROLE OF PSYCHOLOGICAL TESTS IN EDUCATION IN SOUTH AFRICA: ISSUES, CONTROVERSIES AND BENEFITS

K. OWEN

Human Sciences Research Council

Group: Human Resources

© Human Sciences Research Council 1998
All rights reserved

No part of this publication may be reproduced or transmitted in any form of by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without permission in writing from the publisher.

ISBN

K. Owen, D. Litt. et Phil., Chief Research Specialist

Division for Test Development
Group: Human Resources
Executive Director: Dr Sunette van der Walt

Published by:

Human Sciences Research Council
134 Pretorius Street
PRETORIA
0002

ACKNOWLEDGEMENT

The author is indebted to three reviewers for their valuable comments on an earlier draft of this document. Special thanks are thus due to

Prof. C. Plug, Department of Psychology, University of South Africa — for providing a list of corrections and queries;

Mr A.J.J. Brownell, former head of the Educational-Psychological Support Services of the former Natal Education Department — for describing the changed educational scene and the role of testing in the new dispensation;

Prof. M. Skuy, Professor and Head of the Division of Specialised Education, University of the Witwatersrand — for drawing my attention to the importance of dynamic assessment.

CONTENTS

		Page
1.	**INTRODUCTION**	**1**
2.	**MEASUREMENT AND EVALUATION IN PSYCHOLOGY AND EDUCATION**	**12**
2.1	Introduction	12
2.2	Psychological measurement and evaluation	13
2.3	Educational measurement and evaluation	19
2.4	Conclusion	20
3.	**HISTORY OF THE DEVELOPMENT OF PSYCHOLOGICAL TESTS**	**22**
3.1	Intelligence tests	22
3.2	Aptitude tests	25
3.3	Personality tests and questionnaires	34
4.	**APPROACHES TO THE ASSESSMENT OF COGNITIVE DEVELOPMENT**	**40**
4.1	Psychometric approach	40
4.2	Piagetian approach	41
4.3	Soviet approach	44
4.4	Neuropsychologically (biological) based instruments	46
4.5	Dynamic assessment	47
4.6	Conclusion	51
5.	**PSYCHOLOGICAL TESTING: CRITICISMS, ISSUES AND CONTROVERSIES**	**53**
5.1	Criticisms	53
5.2	Test bias	59
6.	**CULTURE AND TESTING**	**69**
6.1	Introduction	69
6.2	What is culture?	70
6.3	The influence of culture on test performance	70
6.4	Possible solutions to the problem of cultural influence on testing	74

		Page
6.4.1	Correction for imperfect prediction	75
6.4.2	Bonus points	75
6.4.3	Within-group norming	76
6.4.4	Top-down selection from separate lists	76
6.4.5	Separate cutoffs	76
6.4.6	Sliding bands	77
6.4.7	A Eurocentric versus an Afrocentric approach to testing	78
7.	**THE ROLE OF PSYCHOLOGICAL TESTS IN SOUTH AFRICAN SCHOOLS**	**86**
7.1	Classification of psychological tests	86
7.2	Test development in South Africa	87
7.3	Cognitive tests developed by the HSRC	89
7.3.1	Individual intelligence scales	89
7.3.2	Group intelligence tests	91
7.3.3	Aptitude and proficiency tests	95
7.4	Affective measures developed/adapted by the HSRC	99
7.4.1	Personality tests and questionnaires	99
7.4.2	Interest questionnaires	101
7.5	Summary	106
8.	**PSYCHOLOGICAL TESTING IN SOUTH AFRICA: END OF THE ROAD OR A NEW BEGINNING?**	**108**
	REFERENCES	**114**

1. INTRODUCTION

It is quite natural at this stage (1997) in the transformation of South Africa that questions about the practice of psychological testing in schools and industry should be raised by the new government officials in various departments and by others. There is a widespread perception that South African psychologists were largely responsible for devising employment instruments that were used to screen out blacks from the workplace in general and higher-paying jobs in particular. The main argument against these instruments (tests) is that they are a Western invention, culturally bound, biased and thus inappropriate to indigenous groups; further, that the constructs measured by these tests and the concepts on which they are based, e.g. aptitude, ability and intelligence, are a European and American middle-class invention and inappropriate in an African context. Fred Zindi (1995) expresses the African perspective thus:

> In the past, a person who exhibited good hunting skills or knew how to look after his immediate and extended family, was a proficient in story-telling, was regarded as intelligent in any African society. With the arrival of the white man in Africa and the resultant aspirations by most urban Africans towards Western technology and intellectual fashions, intelligent behaviour is now being regarded as the ability to solve mathematical problems, exhibiting verbal skills in one of the major European colonial languages and displaying social competence. There is no doubt that these are Western values. ... Western intelligence seems to omit activities which are valued as intelligent behaviour by Africans.

Resistance by blacks to the use of psychological tests has its roots in the USA. In 1975 Jackson (MacKenzie 1981: 234), then President of the Association of Black Psychologists, said that psychological testing "historically has been a quasi scientific tool in the perpetuation of racism on all levels of social and economic importance ... and tests have prevented blacks from gaining access to education, jobs, and housing." After the Black Psychologists Manifesto flatly stated (in 1968)

that psychological tests were intrinsically biased, group intelligence tests were banned in schools in New York, California and Washington DC (MacKenzie 1981: 234). According to this author the evidence for bias in tests takes many forms: "Most persuasive is evidence of mean differences in test scores between minority and majority groups almost always favouring the majority group. Most widely reported is the difference of approximately I S.D. (one standard deviation) between mean intelligence test scores for blacks and whites in the U.S.A." Other factors which have been put forward as evidence of bias stem from the informal content analysis of selected test items, the fact that psychologists and test developers belong almost exclusively to the urban middle class and the belief that minority children in the USA are less experienced with tests than other children and less motivated to do well in tests.

Another scathing attack on intelligence testing comes from Pamela Zappardino (1995). In the abstract of her paper she states the following:

> Stephen Jay Gould points out in *The Mismeasure of Man* (1981), "Science, since people must do it, is a socially embedded activity. It progressed by hunch, vision and intuition". The legacy of the traditional construct of intelligence and its measurement through intelligence quotient (IQ) tests has not been educational improvement. Its legacy in the classroom has most often been the denial of educational opportunity in the guise of cognitive ability grouping. *IQ testing has promoted racism through the placement of students* (emphasis added). The modern construct of intelligence has been narrow, ignoring the many types of intelligences that exist in people. Human ability has been modeled in a manner that has caused harm to many and at great cost in terms of resources, wasted opportunity, and divisiveness. Intelligence tests are actually constructed to produce a bell-shaped curve in which 50 % of test takers are required to score below average. The reasonableness of this process is seldom questioned despite the lack of evidence that intelligence is actually distributed in this way among humans. The truth being

> sought has not been found, and as Frankenstein came to realize, a very long experiment has gone wrong. It is time to give up faith in the numbers generated by testing and to acknowledge intelligence as something other than a straight line, as a construct more resembling a tangled bush than a ladder.

An interesting aspect of the criticism levelled against psychological tests is that it is mainly - or virtually exclusively - intelligence tests that are targeted, i.e. only one type of instrument from a vast array of psychological instruments is singled out for scorn. There is of course a reason for this which will be discussed later in this document.

The question can rightly be asked: Is testing at all necessary? If there were no differences between individuals as far as human attributes are concerned, testing would of course be unnecessary (and neither would differential psychology exist). Dorothy Adkins (1974: 5) puts it thus:

> If all students in a course of instruction had identical aptitudes, interests, health, motivation, and other personality characteristics and if they had been subjected to the same environmental forces, no differences among them would be revealed either at the beginning or at the end of the course. The very natures of the human being and the organization of our traditional educational system, however, ensure that students will differ in relevant characteristics before and after exposure to uniform segments of subject matter. *If individuals did not vary, the field of testing never would have developed* (emphasis added). Faced with differences in abilities, educators and psychologists became interested in how to measure them and in what types of recommendations reasonably could be made upon the basis of these measurements. Individuals do differ markedly in their learning, as reflected in performance, after exposure to a uniform course of instruction that is presented regularly.

One way to measure learner progress in academic areas is by means of

standardized achievement tests. E.L. Thorndike (1874-1949) was a pioneer in developing standardized tests at the beginning of the twentieth century. He believed that if something (such as academic achievement) existed, it existed in some amount and could therefore be measured (Ediger 1994: 169-170). Many educators seem to believe that, since standardized tests are used to measure student achievement, the results are objective. It must be pointed out, however, that this is not necessarily the case, because subjectivity and judgment are involved when determining which items should be included in the test. In spite of these and other limitations, Bali *et al.* (1984) firmly believe that ability and aptitude tests can contribute to the solution of educational problems in developing countries. The use of test results in addition to school grades may offer the possibility of achieving a better and fairer distribution of educational opportunities.

Psychological or standardized testing has many shortcomings (which will be touched upon later). The psychometric approach in particular has significant limitations when used with students of different races, cultures and languages (Hoy & Gregg 1994: 159). However, the negative aspects of testing should be carefully weighed against that which is to be gained by testing. In addition, in evaluating criticism of tests and test items, one should bear in mind the words of Adkins (1974: 4) in this regard:

> Some individuals seem to make a hobby or even a second career out of noting trivial flaws in test items. A recurring theme of such critics is that the really knowledgeable or creative person who takes an aptitude test frequently or regularly will think of nuances of interpretation that lead to answers other than those keyed as correct. Such a criticism is occasionally justifiable, but in the long run it applies to so few items as to be insignificant in affecting important decisions made on the basis of a test - decisions that usually should and do take into account much other data.

Why use psychological tests? Psychological tests provide information about behaviour - usually typical behaviour. This kind of information is of importance to

the individuals concerned and to parents, teachers, psychologists or employers. The same or better information could possibly be obtained by having the person/testee observed by a highly trained expert over an extended period of time, but this is usually impractical, impossible or exorbitantly expensive. Tests can make information available to trained and qualified teachers and psychologists in such a way that appropriate decisions can be made more often than would be the case without the information. As tests are merely samples of behaviour, the generalization of results to the behaviour outside the test situation implies statements of probability rather than certainty. The beneficiaries of testing are those who are enabled to take appropriate decisions more often than they would have been able to without the test. Testees themselves are more knowledgeable about their likelihood of success in certain endeavours; teachers understand more about the attitudes and abilities of their students; psychologists are better able to predict behaviour in related contexts.

The beneficial role of psychological and educational tests in the educational situation can hardly be overestimated. Initial evaluation, for example, is especially helpful in enhancing the aims of instruction (Bouwer 1993). Assessment of the learner's standard of performance at the beginning of a particular course can be used as an indication of the level at which the instruction should commence. Ausubel (1968) puts this succinctly: "If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly!"

Test results are almost indispensable for identifying those students in a class who may require special attention because of learning difficulties. If these difficulties can be timeously identified, the problems can often be solved by appropriate remedial teaching. This aim cannot be accomplished without the use of diagnostic tests. Psychological tests are intended to measure or evaluate certain specific aspects of an individual's cognitive (intellectual) abilities, psychomotor abilities and/or personality traits. The information gained in this way may be used to advise parents and teachers on issues such as:

- school readiness, i.e. whether to send a child to school before compulsory school-going age
- the type of school and curriculum best suited to realize the pupil's full potential
- factors possibly involved in the pupil's poor performance or other behaviour problems
- appropriate remedial programmes for a pupil with learning problems
- deciding upon a special educational programme for a child
- suitable subject choices at school or career guidance and counselling

It should be borne in mind that in all these instances psychological tests give no answers, but merely provide relevant information so that sounder conclusions may follow.

The utility of standardized tests is often severely hampered by a number of misconceptions regarding their use and interpretation. These misconceptions are undoubtedly the source of much of the criticism of tests. In this regard Dyer (Van der Westhuizen 1979: 27) said some thirty years ago that: "Tests *could* be a blessing to education if only teachers and counselors and educational administrators would divest themselves of a number of misconceptions about what tests can and cannot do and would learn to use test results more cautiously and creatively in the educational process."

A major misconception, according to Dyer, is the view that intelligence tests measure "inherent ability", as if it were a quality which one possesses and which remains unchanged throughout life. It cannot be denied, of course, that people have inherent abilities, and there are empirical data to support this assumption. Intelligence tests, however, cannot measure such inherent abilities, nor do they claim to do so. The most that an intelligence test can do is to set a testee certain intellectual tasks and to measure achievement in these tasks. Individuals' ability to complete such tasks successfully has to a large extent been acquired through the experiences they have gone through in their individual worlds. How much individuals learns through experience depends on many factors, such as the clarity

and emotional quality of all the events in their daily lives. It can be assumed, however, that people's innate abilities will to a large extent determine how they interpret and classify their experiences.

A second misconception about standardized psychological tests is the expectation that they will predict with one hundred per cent accuracy, and if they do not meet this expectation, they are rejected as useless. This error is usually based on the misconception that there should be a fixed relationship between a person's test achievement and his or her actual achievement. It would be more meaningful to regard prediction as a statement of probability - human behaviour can never be predicted with absolute certainty.

A third misconception is that achievements in standardized tests are infallible and perfectly reliable. There is, however, a possibility that levels of achievement will vary within a test or between similar tests. Test-users must bear in mind that any test achievement is at best only an estimate of actual ability.

A fourth misconception is that the contents of scholastic achievement tests fully represent the contents of school subjects. No single scholastic test can test all aspects of a particular school subject. On the other hand, there is also the erroneous belief that scholastic tests measure only the pupil's memory for facts. Modern scholastic tests demand that pupils remember the facts, but also that they be able to apply these facts in problematic situations.

A fifth misconception is that personality tests measure constant personality structures. This view can be dangerous, especially when dealing with children. Even if certain personality traits can be clearly described, often comparatively little is known about their stability. In spite of these limitations, personality tests are of importance in school guidance in order to obtain a more complete image of as many personality traits as possible at that stage. Psychological tests do help to form a clearer personality image of the individual.

A sixth misconception is that a series of standardized psychological and scholastic

tests can reveal everything necessary for school guidance. No test or series of tests can provide a complete image of a individual's personality. At best it may estimate levels of development and abilities in various fields.

A seventh misconception is related to the interpretation of interest questionnaires. These questionnaires are used in occupational and study guidance in the senior secondary school phase especially. The interest questionnaire is a very effective instrument for helping pupils or students to get to know themselves more thoroughly. Unfortunately, information obtained by means of these questionnaires is often misinterpreted, as though the questionnaire were able to determine the occupation the person should pursue. This is one of the most dangerous misinterpretations in the field of guidance and counselling. Interest questionnaires are concerned with personal interests only and they do not measure aptitude or any other ability that may be laid down as a requirement for any specific occupation. The interest profile is often interpreted only in terms of the highest fields of interest, instead of relating all fields. Low interests also provide valuable data on a person.

How can psychological tests benefit children in the new educational dispensation in South Africa? If all relevant information for taking sound decisions is available, tests obviously have no role to play. Individual differences are, however, a fact of life and psychological tests show that people differ regarding a variety of characteristics. Knowledge of self will empower the individual to make better informed decisions and to embark on courses of action well aware of the implications of his or her actions. In this way psychological tests can facilitate a more rational and responsible lifestyle. To mention the use of tests in just one educational sphere: it is simply not imaginable that mentally retarded children can be mainstreamed in classes of 40+ and that all the children in that class will receive adequate opportunities for growth. Placement in special education will present a challenge and will involve a large arbitrary and subjective element if the kind of high quality information made available by certain standardized psychological tests is not available.

Searching for equity is imperative in a society that believes that all people are of equal value and ought to be treated as such. It is expected that the new education dispensation in South Africa will accommodate all pupils in one education system, but within this system there will be adequate room for diversity. Diversity may be accommodated in schools that are to a large degree monocultural, as well as in schools in which multicultural education is implemented purposefully at all levels. Whatever the structural differences, important values held by all schools and transmitted to pupils will probably be respect for the individual and respect for cultural differences. The role of psychological tests in schools in the new education dispensation where diverse cultures are to be accommodated should not be very different from the role of psychological tests in the previous education system. After all, it is not news to psychologists that behaviour is influenced by context, and all behaviour should be understood against the backdrop of the context in which it occurs.

In a multicultural context professional judgement may be expected to play a much larger part in the interpretation of test scores than it does in a monocultural context. The cultural environment of each individual and its influence on test scores and expected behaviour have to be borne in mind when test scores are interpreted. It may be necessary to use tests designed for a particular cultural group: here one could think of an individual intelligence test that should preferably be applied in the mother tongue at lower age levels. When a single test is used for various cultural groups, the nature and extent of the *bias* that may arise in the case of a particular pupil should be known to the counsellor in order that these may be accommodated in the interpretation of test scores.

The tests used should preferably have been developed for South African children. Where a single test is developed for all cultural groupings, the possibility of bias against certain cultural groupings should be investigated and any such findings should be reported in the test manual in order that interpretations may be adjusted accordingly. Joint norms for all children may be augmented by cultural norms and local norms, thus rendering raw scores more meaningful for the individual. When the development of a single instrument for all groupings is not possible or practical,

different instruments should be developed for different cultural groups. One may of course end up with a plethora of tests rendering test scores that can *not* be readily compared. This may prove particularly inconvenient in a multicultural setting.

In any educational dispensation, the new dispensation in South Africa included, facilities will be limited by real world constraints. Who will have access to various facilities, for instance education, after the first ten years of compulsory schooling? Who will have access to various training courses at college and university? If decisions in this regard are taken arbitrarily or the only criterion used is membership of a previously disadvantaged community, standardized educational and psychological tests have no place. If, on the other hand, we are creating a fair and just society where certain values are explicit, tests that offer valid and relevant information will be able to make valuable contributions.

It is assumed that we will be living in a society where resources will be optimally used and excellence will be a commonly accepted virtue. In the absence of these values it will not really matter how well something is done. Even though people may differ in abilities and characteristics it will not really be of importance that the people best equipped to do certain jobs do them or, for that matter, that they do jobs that are well suited to them. Under these circumstances special abilities will not be recognized and potentialities will not be realized. On the other hand, judicious selection and well considered choices will optimize human development.

It is evident from the above that there are many conflicting views on testing. Tests have ardent supporters and equally fervent opponents. These conflicting views are captured very aptly by Hopkins and Stanley (Ediger 1994: 5) who, referring to the *paradox* of testing, write, "Many people are opposed to measurement and evaluation, yet at the same time favor excellence, which is facilitated by and can be identified only through measurement and evaluation."

The main purpose of this monograph is to

- stimulate debate on the role and use of standardized tests in South Africa;

- assist decision makers, especially in the education field, in the use of psychological instruments;

- convey as much information as possible about tests and measurement without confusing readers who are not measurement experts.

Matters that will be touched upon include

- measurement and evaluation in psychology and education,

- the history of the development of psychological tests (intelligence/aptitude/personality tests and questionnaires),

- approaches to the assessment of cognitive development and abilities,

- criticisms and controversies surrounding tests and testing (e.g. test bias),

- the role of psychological tests in South African schools with the emphasis on what the HSRC has to offer in this regard, and lastly

- some thoughts on the future of testing in South Africa.

2. MEASUREMENT AND EVALUATION IN PSYCHOLOGY AND EDUCATION

2.1 INTRODUCTION

Measurement and evaluation have become such a part of our everyday lives that no particular attention is given to them. Just think of all the examinations that are written annually in schools, colleges, universities, etc. Consider the industrial psychologist who has selected a number of applicants for certain posts: How successful was the selection? Or an education department which introduces a new teaching method: How successful is it? These are only a few examples of the role played in society by measurement and/or evaluation.

But what exactly do *measurement* and *evaluation* mean? Measurement is the process of determining, by means of observation and testing, the characteristic features of specific entities and allocating a number, a score or an assessment to the result (Goodwin & Driscoll 1980). Measurement therefore concerns scales, numbers and constructs. The *aim* of measurement is elucidated as follows by Green (1970: 4): "Measurement is concerned with the application of an instrument or instruments to collect data for some specific purpose" and evaluation defined as "the process of subjective appraisal with specific purposes or aims in mind".

Evaluation is a term that has wider implications than measurement and it can be regarded as the process in terms of which the *value* of something is assessed. This often occurs in terms of costs, applicability or effectiveness (Goodwin & Driscoll 1980).

The administration and scoring of a test are regarded as part of the measurement process; findings regarding the score obtained (for instance, whether it is good or bad, depending on the purpose for which it is intended) are regarded as part of the evaluation process. According to Tuckman (1975: 12), evaluation is "a process wherein the parts, processes, or outcomes of a program are examined to see whether they are satisfactory, particularly with reference to the program's stated

objectives, our own expectations, or our own standards of excellence."

Evaluation is applicable to any activity, programme, product or person and usually ends when an assessment has been made. Evaluation therefore involves values, needs, measurement and criteria. Because evaluations are usually based on data, measurement is an extremely important facet of the evaluation process. The credibility of evaluation is therefore also closely related to the *quality* of measurement.

2.2 PSYCHOLOGICAL MEASUREMENT AND EVALUATION

What is the value of psychological measurement? "We consider psychological measurement to be an essential component of all kinds of counseling. We value it both as a source of diagnostic information about clients and as a stimulus to self-exploration and self-understanding for clients" (Seligman 1994: 63).

What is a psychological test? According to Anastasi (1976: 23) it is "essentially an objective and standardized measure of a sample of behavior".

The definition given by Russel and Cronbach (1958: 217-223) agrees with this: "Psychological tests are nothing more than careful observations of actual performance under standard conditions."

The term "careful" implies that the sampling of the performance or behaviour and the obtaining of a record of it are systematic and objective enough for different observers to make reasonably comparable findings.

Testing therefore involves a measuring instrument by means of which a person displays her or his behaviour by answering questions or solving problems. *Names* that are given to such measuring instruments include questionnaire, attitude scale and aptitude test. One of the most important functions of psychological tests is to measure inter- and intra-individual differences, in other words differences between individuals as well as differences in the various individuals themselves.

One of the first areas in which psychological measurement played a role was in the identification of mentally retarded people, and the determining of intellectual handicaps currently remains an important function of certain types of psychological tests.

Psychological tests can, however, be used for many other purposes, such as the determination of individual differences in general intelligence, specific aptitudes and non-cognitive personality traits. Tests have also been used for some time in psychological, educational, cultural, sociological and occupational situations. Psychological tests, particularly general intelligence and aptitude tests, are used on a large scale in education - from first grade to university level - for classification, selection and planning.

Tests are used to determine and analyse intellectual abilities or personality traits in order to provide school and vocational guidance, to place pupils in special classes for gifted or mentally retarded children, to identify weaknesses such as reading disabilities with a view to offering remedial teaching and to determine and remove intellectual or other causes of behaviour problems at school.

In clinics, tests are mainly used in respect of problems related to learning or progress at school, attitudes, interpersonal relations, emotional disturbances, juvenile delinquency and other behaviour disorders.

According to Mehrens and Lehman (1973), psychological measurement has four aims:

(1) Teaching: This aim is closely linked to the learning process in the sense that the learning of new behavioural patterns by students, clients or employees is monitored on an ongoing basis. This enables the person who is responsible for guidance (such as the teacher, industrial psychologist or counselling psychologist) to encourage desirable and discourage undesirable behavioural patterns.

(2) Counselling: People continually experience a need for counselling in respect of educational programmes, occupational choices and personal problems. Aptitude tests, interest questionnaires, personality questionnaires and achievement tests can be used successfully in support of such counselling.

(3) Administration: This aim has a special bearing on the selection, placement and classification of employees and even pupils. The use of psychological measuring instruments facilitates the appointment of personnel and enables training officers and teachers to devise new training programmes and adapt existing ones.

(4) Research: This aim is fundamental to all three of the above aims as decisions are often based on research findings.

In addition to their utility in solving a wide variety of practical problems, psychological tests have another very important use as measuring instruments in basic research. Almost all of the problems encountered in differential psychology, such as the nature and extent of specific individual differences, the measurement of group differences and the biological and cultural factors that have a bearing on certain behaviour differences, require testing procedures as a means of obtaining information.

The general utility of psychological tests can be summed up as follows (Tuckman 1975: 7-9):

- to lend objectivity to our observations
- to elicit behaviour under relatively controlled circumstances
- to sample the behaviour people are capable of
- to measure the progress made with regard to set objectives or standards
- to give insight into aspects of human beings that are not visibly observable
- to trace characteristics and components of behaviour
- to predict future behaviour
- to provide information for feedback and decision making

Measurement in the social sciences is therefore a controlled and relatively objective procedure by means of which the behaviour a person is capable of can be determined and assessed against a norm or specific standards. This method can facilitate the feedback of information to testees, the diagnosis of learning disabilities and other weaknesses, the tracing of special skills, knowledge and creativity, the discovery of character, temperament, values, interests and much more. However, the success of measurement depends on how well the measuring instrument has been compiled, how well it is administered and how skilful the user is in interpreting measurement results.

It must further be borne in mind, as Zeidner and Most (1992) point out, that psychological testing is based on the assumption that decisions made in educational, vocational, clinical and other settings involve a certain amount of uncertainty or risk with respect to outcomes. Decisions should therefore be based on information as reliable and comprehensive as possible. To facilitate decision making, tests are designed to provide objective and reliable information to serve as inputs. In general, test results can be used to assist clients and psychologists in making decisions and in choosing optimal courses of action.

A very important matter in testing, one that is often overlooked by psychologists, is that examinees taking a particular test should be similar in cultural, educational and social background and experiences to those on whom the test has been standardized and the test norms based. If the testee or group differs from the standardization sample, the use of the norms for evaluating current performance or prediction may be inappropriate. According to Zeidner and Most (1992: 22), "if a test shows different levels of accuracy in assessing the target construct or in predicting a criterion score as a function of subcultural or gender group membership, then that test may not be appropriate for all cultures or genders".

Psychological measurement also has its problems. Owing to the complex nature of the human personality the conditions established by Thorndike and Hagen (1969) with respect to measurement can seldom be met. According to these writers, all measure-ment should comply with the following three requirements:

(1) The attribute being measured must be clearly identified and defined.

(2) A decision must be made regarding how the specific attribute or characteristic can be observed.

(3) Procedures must be determined for converting observation into quantitative data.

As far as the first requirement is concerned, the concept *intelligence* is a good example of the problems that may arise. Most people can provide a general description of *intelligent* behaviour, but experience problems as soon as a more particular definition is required. Difficulty in formulating an exact definition of a human characteristic or quality is experienced with regard to many psychological and educational concepts.

The second requirement - to determine methods for isolating the characteristic and making it observable - also presents problems. The definition of a characteristic and the method by which it is optimally isolated for observation are often closely related and together they form an *operational definition*. This means that the methods used to isolate an attribute can actually be regarded as defining the attribute or characteristic concerned. The definition used will in turn suggest a relevant and suitable method for revealing the characteristic. For instance, the aim of the methodology for standardizing tests is to develop "instruments and procedures for eliciting in a standard way and under uniform conditions, the behaviors that serve as indicators of the relevant attributes of persons" (Thorndike & Hagen 1969: 12).

Numerous problems can therefore develop as a result of the definition of concepts and the specific definitions that are used.

Problems also develop when psychological measurement has to comply with the third requirement - that observations of behaviour be converted to quantitative data. Most psychological and educational measurements occur at the ordinal and even the nominal level. The difference between rankings and the numbers that are deduced from them is usually the highest possible level of objective measurement

of human behaviour. The measurement of a pupil's IQ merely means that his or her place in the rank order has been determined with respect to other pupils of the same age.

In spite of the fact that psychological measurement cannot comply with all the requirements of interval measurement, the objective measurement of certain non-cognitive human characteristics has made a considerable contribution towards relatively valid description.

It should also be kept in mind that the human being as a totality cannot be measured, but that the human being's psychological composition alone is measured. Reference to momentary psychological *measurement* thus implies the measurement of no more than the momentary condition of an aspect of the quality or characteristic concerned.

Measurement in psychology is closely related to *evaluation*. A single score in the form of a test result contributes very little to a personal description, for instance, unless it is evaluated in the light of other information concerning the individual, whether objective or subjective. Ahmann and Glock (1959: 11) say, "In the last analysis measurement is only a part, although a very substantial part, of evaluation. It provides information upon which an evaluation can be based."

Psychological evaluation can therefore be defined as a process in which the psychologist uses information obtained from a wide variety of sources in order to make a value judgment. The data can be obtained with the aid of psychological tests and other techniques that do not necessarily provide quantitative results. Standardized measuring instruments need not be included here either, although they contribute a large measure of objectivity to the evaluation.

Evaluation is more comprehensive than measurement because value judgments are involved. Subsequent to the measurement of aspects of an individual's personality, intellectual abilities or the like, the results are expressed in standards or norms for the age group or in terms of the purpose for which the measurement

was undertaken. Practically speaking, however, measurement and evaluation cannot be separated because in most cases evaluation occurs or should occur together with the measuring procedure.

2.3 EDUCATIONAL MEASUREMENT AND EVALUATION

If the definition in 2.1 is applied to educational results, educational measurement can be defined as the allocation of scores to the results of instruction and/or learning at school. The instrument developed for this purpose is called an achievement test. The results of learning at school are usually inferred from the pupils' understanding of some or other measure of knowledge or from their proficiency in certain skills. In other words educational measurement involves the evaluation of the achievement of pupils in some field or other - usually a school subject.

Educational evaluation is a wider concept. First of all, educational evaluation can be based on either quantitative or qualitative data and necessarily involves a value judgment. Stanley and Hopkins (1972: 3) distinguish between educational measure-ment and evaluation in this way: "We consider the construction, administration, and scoring of tests as the measurement process. Interpreting such scores - saying whether they are good or bad for a specific purpose - is evaluation."

As implied in the definition and explanation above, evaluation in education involves far more than the traditional testing, examination, classification and promotion of pupils and reporting back to parents that are such an important part of the regular school programme (Tyler 1966: 18-19). Evaluation must be regarded as an integral part of the educational process since education and evaluation are bound together in an unending cycle of change. It is quite normal for the results of evaluation to lead to the reformulation of certain educational objectives and in turn changes in the educational programme. The latter again transform the evaluation programme, and so the cycle repeats itself in a process of progressive improvement. Evaluation plays its role in education as "a recurring process

involving the formulation of objectives, their clearer definition, plans to study students' reactions in the light of these objectives, continued efforts to interpret the results of such appraisals in terms which throw helpful light on the educational program and on the individual student" (Tyler 1966: 25).

The role of evaluation in education can also be described in relation to three methods of evaluation (initial, formative and summative evaluation) that are in turn identified by the way in which evaluation affects the various stages of the teaching/learning process. One can start by emphasizing the need for *initial evaluation*, in other words that "the teacher must be able to diagnose the relevant characteristics of his learners at the time they enter the course or program" (Bloom, Hastings & Madaus 1971: 15).

It would obviously be presumptuous of a teacher to start a work programme without first ascertaining the standard of a new group of pupils in the subject concerned in order to ensure that the new work can build on the foundation of existing knowledge.

A second phase of evaluation takes place during the course of the teaching/learning process, namely *formative evaluation*. The aim here is primarily diagnostic. Both pupils and teacher are provided with feedback on the effectiveness of learning and teaching in each stage of the instructional process. Such evaluation forms the basis of a system of quality control (Bloom *et al.* 1971: 8).

Lastly it is of course necessary to conclude a programme or course of instruction with an evaluation of the measure of proficiency reached in the whole learning unit before commencing with a new one. This is called *summative evaluation* and it forms the basis of the traditional progress reports provided to parents.

2.4 CONCLUSION

In all contexts of education and training, measurement and evaluation of the

learner's progress form an essential part of instruction. Test results provide information by which to assess the standard of the instruction and make decisions about the learner's future. Without testing, instruction would lose much of its intrinsic motivation and could become superficial. The quality of education for a particular learner can be considerably enhanced by the judicious application of information derived from psychological tests. The psychologist has to assume the responsibility for using a test appropriate to the particular circumstances. The test publisher in turn has the obligation to provide relevant and accurate information about the reliability and validity of the tests published and about the steps taken to eliminate test bias with regard to aspects such as language, gender, culture and socio-economic status.

3. HISTORY OF THE DEVELOPMENT OF PSYCHOLOGICAL TESTS

3.1 INTELLIGENCE TESTS

Tests of many different kinds have proliferated in the past seventy-five years. During this period the use of standardized tests as diagnostic and predictive instruments in elementary and secondary schools has increased markedly. Procedures in the USA for the selection and appropriate placement of employees by government, the armed services and industry ordinarily include the use of various kinds of measuring instruments (Adkins 1974).

The development of cognitive tests has its origin in the work of Sir Francis Galton (1822-1911), a cousin of Charles Darwin, and of James McKeen Cattell (1860-1944) towards the end of the 19th century. Galton was interested in reaction times, while McKeen Cattell was interested in finding out what underlying basic abilities were projected into actual performances - the methods used at that time (1901) did not, however, permit a solution (Cattell 1983). Galton's work in 1883 was fortunately more fruitful in that it led "to the recognition that human traits tended to be *normally* distributed and it led also to the development of the correlation coefficient for determining how much any two abilities are related" (Cattell 1983: 227). In 1879, Wilhelm Wundt (1832-1920) established his psychometric laboratory in Leipzig in Germany to study mental events by introspection. McKeen Cattell studied with Wundt as well as Galton and imported this knowledge to the United States (Li 1996: 2).

At the turn of the 20th century, interest in intelligence grew rapidly. Alfred Binet (1857-1911) and Theodore Simon (1873-1961) were the first to develop an intelligence test for French children. Binet was instructed by the school authorities in Paris to develop such a test because the schools were concerned to separated out those who merely performed poorly in school from those who were mentally retarded.

Charles Spearman (1863-1945), a British psychologist, perfected statistical

techniques to measure intelligence and in 1904 also invented factor analysis to treat mental test scores. He postulated a two-factor theory, a simple general factor, "g", to denote general intelligence, and several independent specific factors, "s", to represent interrelations among the tests. Lewis Terman (1877-1956) revised Binet's scales in 1916 to create the Stanford-Binet Scales for American children. Today this test is still one of the most popular IQ tests in America and the latest revision was published in 1986. It was in this test that the term 'intelligence quotient' (IQ), or ratio between mental age and chronological age, was used for the first time[1] (Anastasi 1976). According to Cattell (1983), it was Stern, in Germany, who pointed out that if one divided what he called the mental age by the actual age of a child one attained what he called intelligence quotient and that this intelligence quotient remained essentially constant over the years of the child's development. Unfortunately this tended "to be interpreted as meaning that the I.Q. measured a relatively innate general ability, but in a strictly logical approach one could account for it both as due to heredity and as due to a uniformity of the lives of most children in relation to school experience, as they grew up" (Cattell 1983: 229). Other individual scales for children include the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and the Wechsler Intelligence Scale for Children (WISC). The first individually applied intelligence scale for adults was prepared by David Wechsler. This scale, known as the Wechsler-Bellevue Intelligence Scale, was published in 1939; it was later supplanted by the Wechsler Adult Intelligence Scale (WAIS).

The Binet tests, and the revisions that followed, are all individual scales, i.e. they can be administered to only one individual at a time. The need for *group* testing arose when the United States entered World War I and a great many new recruits had to be rapidly classified. One of the first group intelligence tests was compiled by Arthur S. Otis and out of this followed the Army Alpha and Army Beta tests. After the war, these tests were turned over for civilian use, a step that had far-reaching consequences. According to Anastasi (1976: 13), these tests did not only serve as models for new group intelligence tests but

---

[1] This type of scale has long since been replaced by a deviation IQ scale with a mean of 100 and a standard deviation of 15 that has nothing to do with a ratio between mental age and chronological age; the term "IQ" has, however, been retained.

> The testing movement underwent a tremendous spurt of growth. Soon group intelligence tests were being devised for all ages and types of persons, from preschool children to graduate students. Large-scale testing programs, previously impossible, were now being launched with zestful optimism. Because group tests were designed as mass testing instruments, they not only permitted the simultaneous examination of large groups but also simplified the instructions and administration procedures so as to demand a minimum of training on the part of the examiner. Schoolteachers began to give intelligence tests to their classes. College students were routinely examined prior to admission. Extensive studies of special adult groups, such as prisoners, were undertaken. And soon the general public became IQ-conscious. The application of such group intelligence tests far outran their technical improvement. That the tests were still crude instruments was often forgotten in the rush of gathering scores and drawing practical conclusions from the results. When the tests failed to meet unwarranted expectations, skepticism and hostility toward all testing often resulted. Thus, the testing boom of the twenties, based on the indiscriminate use of tests, may have done as much to retard as to advance the progress of psychological testing.

Probably one of the best known group intelligence tests, is the Otis-Lennon School Ability Test (OLSAT), which was published in 1918. In recent years a number of nonverbal tests of intelligence have been published in reaction to criticisms of intelligence tests being culturally biased (Seligman 1994). These tests are especially useful in testing people with language difficulties. Examples of nonverbal tests include the Test of Nonverbal Intelligence (TONI), Cattell's Culture-fair Intelligence Test and Raven's Progressive Matrices. Although these tests typically do not rely heavily on school learning, they may provide misleading information in respect of people who are aspiring to educational or occupational goals where a high level of verbal ability is of importance (Seligman 1994).

3.2 APTITUDE TESTS

Multiple aptitude test batteries represent a relatively late development in the testing field and nearly all have appeared since 1945 (Anastasi 1990: 15). The development of these batteries can be attributed mainly to the different selection programmes used by the defence force and large industries in the United States, as well as to a realization of the inadequacy of intelligence tests to explain intra-individual differences (i.e. the differences between the abilities of the same person). Intelligence tests were originally constructed with the objective of measuring a wide variety of functions in order to estimate the *general intellectual level* of the individual. Although these tests measured certain key functions, it gradually became evident that they were of limited value because they did not cover other important functions such as mechanical ability. Multiple aptitude tests (e.g. the HSRC aptitude tests), in contrast to general aptitude tests (intelligence tests), have a differential approach to the measurement of aptitude. Such an instrument does not provide a single or total score such as an IQ, but rather a set of scores in respect of different aptitudes. With the help of these scores an intellectual profile showing the individual's characteristic strong and weak points can be drawn.

In order to place aptitude and the ability or abilities at issue in perspective, it is necessary to refer to some of the theories on the structure of intellectual abilities. The theories of Spearman, Vernon, Thurstone and Guilford will therefore be briefly discussed.

(1) Spearman's two-factor theory

Spearman's two-factor theory was mentioned in paragraph 3.1. This theory on the structure of intellectual abilities was the first that was based on a *statistical* analysis of test scores (Anastasi 1990: 381).

According to this theory, all intellectual activities share a common factor, called the general factor, also known as "g", and a specific factor, "s", which is unique to the

particular test. Apart from the common factor g, there are therefore just as many s factors as there are different activities or tests. A positive correlation between any two activities (tests) is accordingly ascribed to the g factor, and the higher the correlation the greater is the g "saturation" of the tests. The presence of specific factors, i.e. factors unique to the particular activities or tests, however, tends to lower the correlation mentioned between the activities.

Although two types of factors, general and specific, are postulated by the theory, only one factor, the general factor g, is responsible for the correlation between activities. Consequently Anastasi (1990: 381) correctly observes that Spearman's theory should actually be called the one-factor theory - however, the original term has become so widely accepted that it cannot be changed now.

In about 1925 Spearman, in collaboration with Holzinger and others, began to investigate the specific factors (s), or *group factors*, as they would later be known (Carroll 1993: 637). The model that emerged from this co-operative research was called the *bifactor model* by Holzinger, and was in essence a two-strata model with g in the higher stratum and a variety of group factors - such as arithmetic, mechanical and linguistic abilities - in the lower stratum.

From the two-factor theory it follows that psychological measurement should endeavour to measure the amount of g in an individual. Spearman therefore proposes that intelligence tests consisting of heterogeneous items should be replaced by a single test that measures mainly g. Tests that best meet this requirement are, according to him, tests that are concerned with abstract relationships. Such tests include Raven's Progressive Matrices and Cattell's Culture Fair Intelligence Tests (Anastasi 1990: 382).

In the 1960s Cattell (1983: 231) found that Spearman's g split into two distinct gs which have been called g_f, fluid intelligence, and g_c, crystallized intelligence. The main difference between the two kinds of intelligence is that fluid intelligence is involved in tests that have very little cultural content, whereas crystallized intelligence involves abilities that have obviously been acquired, e.g. verbal and

numerical ability, social skills, and so on.

(2) Vernon's hierarchical model of abilities

Vernon was a colleague of Spearman and his structure of intellectual abilities is reminiscent of a family tree: he places Spearman's g factor at the top of the hierarchy and on the next level two broad group factors which he calls verbal-educational (v:ed) and practical-mechanical (k:m) aptitudes. These two main group factors are in turn divided into a number of smaller group factors. For example, the verbal-educational factor is divided into verbal and numeric subfactors, while the practical-mechanical factor is divided into subfactors such as spatial and mechanical information.

According to Vernon (Carroll 1993: 60), it is an oversimplification to represent this model in the form of a tree as is so often done in textbooks (e.g. in Anastasi 1976: 375); in reality the relationship between the various factors is far more complex. For instance, the general factor (g) dominates the higher-order factors (v:ed and k:m), which in turn dominate a number of smaller group factors, while the latter dominate a variety of very narrow and specific factors. Factors that are dominated by the v:ed group factor include logical reasoning and verbal, numeric and fluency abilities, while the k:m group factor is dominant in respect of factors concerning technical subjects, mechanical information, spatial ability, drawing, handwork, reaction time and psychomotor co-ordination. However, according to Vernon, the g factor is the most important: "most of the variance of human abilities in daily life is attributable to g" (Carroll 1993: 60). People with a high g factor tend generally to do better in most areas (for example, in virtually all the tests of an aptitude test battery - the task of a counselling psychologist, who usually makes specific recommendations on the basis of intra-individual differences, is thus complicated in this instance) than those with a low g factor. On the other hand, there are also cases where people with a relatively low g factor emerge as leaders in various walks of life, such as the sciences, the arts, politics, and so on. Such outstanding achievements can probably be ascribed to strong group factors as well as certain

personality characteristics, motivation and interests.

As far as the "validity" of the model is concerned, empirical data suggest that it probably is valid. From factor analyses conducted by different researchers a general factor has, for example, often been found as well as group factors that manifest the typical characteristics postulated by Vernon.

(3) Thurstone's multiple factor theory

In contrast with Spearman's general factor, Thurstone proposes a number of group factors which he calls *primary mental abilities*. These factors are the following (Anastasi 1990: 383-384):

- Verbal Comprehension (V): found in tests such as reading comprehension, verbal analogies, verbal reasoning and vocabulary.
- General Reasoning (I): encountered in tests for inductive reasoning, i.e. tests where the testee must find a rule, for example number series.
- Word Fluency (W): occurs in tests that require the naming of specific kinds of words, for example words that begin with be-.
- Memory (M): occurs in tests for rote memory, for example where members of a pair are associated with one another.
- Number (N): represented by the speed and accuracy with which simple arithmetical calculations are carried out.
- Spatial (S): stands in connection with geometric figures and the imaginary manipulation of such figures.
- Perceptual Speed (P): occurs in perceptual tasks, for example when similarities and differences between visual stimuli have to be perceived rapidly and accurately.

Apart from the above seven factors, Thurstone also identified two additional factors which he provisionally called D (Deductive Reasoning) and R (possible Restriction) (Carroll 1993: 54).

Although Thurstone's model proposes seven or more primary abilities but does not make provision for a general factor, Carroll (1993: 638) points out that the model was established as early as 1938 and that it is not necessarily the same model that Thurstone championed in his later years. According to Carroll (1993: 56), Thurstone conceded that there might be a correlation between his primary factors and that Spearman's general factor could well exist.

From this it is clear that the allegedly fundamentally different standpoints of the "British" school on the one hand as represented by Spearman and of the "American" school as represented by Thurstone were in reality not so different. On closer inspection it appears that the differences were relative rather than absolute: Spearman and his followers stressed the general factor and regarded the group or primary factors as less important; in contrast Thurstone and his followers considered the primary abilities to be the most important and the general factor less so. For Thurstone the primary factors were crucial, especially because of the *application* or use thereof in, for example, vocational guidance.

As will be seen later, most aptitude tests follow a differential approach to the measurement of abilities - which is in essence Thurstone's approach.

(4) Guilford's structure of the intellect

Guilford's model of the intellect consists of a boxlike figure with 120 cells made up of three dimensions, *operations*, *contents* and *products*. Each cell is described in terms of the three dimensions and represents at least one factor or ability. Each of the three dimensions in turn consists of a number of categories. In the case of *operations* (what the testee does) there are five categories: cognition, memory, divergent production, convergent production and evaluation. *Contents* refers to the nature of the material or information on which the operations are carried out and consists of four categories: figures, symbols (e.g. letters and numbers), words and behaviour (e.g. information on the person's attitudes, needs, etc.). *Products* concerns the form in which the information is processed by the testee and contains six categories: units, classes, relations, systems, transformations and implications.

The 120 cells of the box are thus formed by the 5 x 4 x 6 categories of operations, contents and products respectively. From the preceding it is clear that Guilford's premise was that any factor or variable (test) has not only one but three aspects, facets or dimensions (operations, contents and products). In other words any factor or test that measures that particular factor requires the testee to carry out the one or other *operation* on a certain type of *content* which results in a certain type of *product*.

Although Guilford's model enjoys reasonably wide acceptance in textbooks on psychological measurement, Carroll (1993: 59) questions the logical validity of the model, particularly the way in which the interactions of the different facets lead to factors. According to Carroll (1993: 638), the fact that the model does not make provision for a general g factor must be ascribed to Guilford's somewhat idiosyncratic methodology.

Against this broad background of theories on the structure of intellectual abilities, specific attention will now be given to *aptitude* and its measurement. An important consideration that should not be lost sight of is that although the concept *aptitude* has its own definitions and terminology, its building blocks, namely abilities, are not in any way different from those that were discussed in the previous theories on intellectual abilities; only the context in which reference is made to the abilities differs. To quote Anastasi (1990: 15) on this point:

> The term 'aptitude test' has been traditionally employed to refer to tests measuring relatively homogeneous and clearly defined segments of ability; the term 'intelligence test' customarily refers to more heterogeneous tests yielding a single global score such as an IQ. Special aptitude tests typically measure a single aptitude. Multiple aptitude batteries measure a number of aptitudes but provide a profile of scores, one for each aptitude.

In the literature on aptitude, related terms such as skill, ability, capacity, talent and

potential are often encountered, and it is therefore essential to know what is meant by each of these terms.

(i) Skill

Skill is behaviour or action at a given moment. If a typist can type 70 words a minute, this score of 70 represents skill. The level of skill can change from time to time.

(ii) Ability

Ability is "the power, at a given time, to perform acts or skills" (Gekoski 1964: 41). Ability is the basis of skill. Ability, just like aptitude, is a hypothetical construct - an abstraction. Ability (power) can be expressed in behaviour (skill) and can also be deduced on the basis of skill. Skill is observed and measured. Arising from this measurement, deductions are made about the level of the ability. A single ability can form the basis of different skills.

(iii) Capacity

Capacity is potential ability, in other words the ability an individual may have at a certain time in the future if optimal development takes place in the meantime.

(iv) Talent

Talent is aptitude at a very high level, in other words the person in whom it is manifested is extremely amenable to learning and instruction up to an unusually high level.

(v) Potential

The *Psigologie-woordeboek* (Plug *et al.* 1986: 28) defines potential as follows:

> Characteristics of a person (or of a matter) that will enable him at a

later stage to reveal behaviour or characteristics of a certain kind. The term is mainly used in respect of characteristics that will make it possible for a person to attain exceptional achievements, for example to achieve success in an occupation, at some point in the future (translation).

From the preceding description it is evident that capacity and potential have largely the same meaning.

But to return to aptitude and its measurement: aptitude can be regarded (Fouché & Verwey 1978: 3) "as the potential which a person has and which enables him to attain a specific level of ability with a given amount of training and/or practice. Aptitudes, together with other personality characteristics such as interest, attitude and motivation as well as training and instruction, will determine the level of skill and proficiency which may be reached".

The term *aptitude* is used here as a synonym for specific mental ability, as opposed to general mental ability, i.e. intelligence. In the light of the results of factor analyses, the term aptitude can also be associated with the concepts *group mental factor* (Vernon's model) and *primary mental ability* (Thurstone's model).

Any test, according to Bingham (1937), is a test of aptitude insofar as the score gives an indication of future potentialities. *Predictive value* is therefore the most characteristic feature of an aptitude test: without it a test is simply *not* an aptitude test. With an aptitude test we wish to determine whether a person *now* has the ability to carry out a certain task in the *future*, if given the necessary training in the intervening period. In other words we wish to determine whether a person has the necessary learning ability in a specific direction to enable him or her to achieve success in that direction if appropriate stimuli are provided. An important proviso regarding the interpretation of inter-individual test score differences is, however, that all the testees should have been exposed to more or less the same experience regarding the characteristics that are measured before the application of the aptitude test. If some testees have a lot of experience in a specific area which can

influence their test scores significantly, the counsellor will have to take this into consideration in the interpretation of their scores. Under such circumstances the test scores could be a reflection of skill rather than aptitude. Only if all the testees have roughly the same experience can any meaningful conclusions be drawn about inter-individual differences (i.e. differences between individuals).

Numerous standardized multiple aptitude tests (batteries) are available locally and overseas. Among the best-known batteries developed in the USA are the Differential Aptitude Tests (DAT) and the General Aptitude Test Battery (GATB). The DAT consists of eight tests: Verbal Reasoning, Numerical Ability, Abstract Reasoning, Clerical Speed and Accuracy, Mechanical Reasoning, Space Relations, Spelling and Language Usage. Although the DAT is not based directly on factor analysis and therefore does not measure pure factors but rather group factors, the compilers of the battery are nevertheless led by the results of factor analytical investigations in their choice of tests and items. The needs in the guidance and educational fields are considered of greater importance in the construction of the tests than the factorial purity of the tests. This point of departure also applies to the HSRC aptitude tests. In contrast to the DAT, the GATB, which was developed by the American Department of Labour for use in the public service, was more squarely based on factor analysis. This battery consists of 12 tests which measure the following nine factors or aptitudes:

G - Intelligence
V - Verbal Aptitude
N - Numerical Aptitude
S - Spatial Aptitude
P - Form Perception
Q - Clerical Perception
K - Motor Co-ordination
F - Finger Dexterity
M - Manual Dexterity

The above descriptions of the DAT and the GATB illustrate the fact that aptitude

tests "traditionally" measure certain abilities, for example reasoning ability (whether through verbal and/or nonverbal material), verbal/language comprehension, numerical ability, spatial ability and perceptual speed. However, aptitude test batteries do not necessarily include tests such as mechanical insight/reasoning, memory and co-ordination. This also further shows that the HSRC aptitude test batteries (see the HSRC Test Catalogue) are in many respects typical of similar batteries that have been developed elsewhere in the world.

Although multiple aptitude tests provide reasonably constant measurements and retesting of an individual seldom leads to an improvement in achievement, the tests do not always differentiate to the desired extent. Under these circumstances it is sometimes extremely difficult to provide individual counselling. A typical example of such a situation is where a person obtains either very low or very high scores in nearly all the tests in the battery and consequently a profile cannot be drawn which shows the characteristic strengths and weaknesses of the person (in other words the profile is not differentiated).

The scores obtained from aptitude tests should be regarded as useful pieces of information that can be used with other information about a person in order to take certain decisions. By "other information" is meant school examination marks, interests and attitudes, study habits, hobbies, human relations, particular likes and dislikes, and so on. It should be remembered that aptitude tests are not the "decision maker" but that they provide important information on the basis of which the pupil or student - in consultation with parents, teacher and counsellor - can reach realistic and judicious decisions on, for example, subject or occupational choices.

3.3 PERSONALITY TESTS AND QUESTIONNAIRES

"Although the term 'personality' is sometimes employed in a broader sense," Anastasi (1990: 523) declares, "in conventional psychometric terminology 'personality tests' are instruments for the measurement of emotional, motivational, interpersonal, and attitudinal characteristics, as distinguished from abilities."

The psychological concept of *personality* differs from the popular understanding of the term. To the layperson some people have a strong, a weak or an attractive personality, and some people even have no personality at all. A person with *no* personality has no charm, for example, or is submissive and plain. To the psychologist there is no such thing as a person without personality, but psychologists have not yet agreed on the exact meaning of the term. Hjelle and Ziegler (1976: 18) maintain that a psychologist's definition of personality depends on the personality theory he or she accepts. According to Stagner (1974: 10), Gordon Allport's definition of personality complies with most of the requirements stipulated by psychologists for such a definition. Allport (1961: 28) defines personality as the dynamic organization within the individual of those psychophysical systems that determine his characteristic behaviour and thoughts.

Allport's definition can lead one to conclude that, as a dynamic organization of systems that has developed on the basis of an infinite number of developmental and genetic influences, personality is unique to every individual and unrepeatable. Many psychologists believe that this uniqueness implies that personality should be studied as an organized whole or *gestalt*. To fragment personality into traits is taboo, because the whole is more than the sum of its parts. To illustrate, water is completely different from the elements of which it consists. According to Stagner this theory is based on an inappropriate analogy. Water is destroyed if its hydrogen and oxygen elements are separated, but it can also be studied in terms of variables or characteristics such as temperature, volume, colour and rate of flow without destroying it. In the same way personality can be divided into variables or traits without harming the unique total image.

According to Semeonoff (1966: 8), the main field of personality study was established only late in the 1930s. Before that time the term personality was used primarily in the description of abnormal phenomena, while variations in *normal* personalities became the chief area of study following the introduction of personality studies. The emphasis thus shifted from description to quantification and experimentation. Measurement, that is accurate measurement, is a prerequisite for scientific quantification and experimentation. The results of

experimental or controlled intervention in human circumstances can be scientifically assessed only in terms of differences in measurements, which are sometimes small but nevertheless statistically and practically significant.

Various forms of personality measurement have existed since the earliest times. However, the use of personality tests to provide measurements or to establish a basis for systematic description is a result of personality study which is, as Semeonoff correctly points out, a fairly recent development in the relatively young science of Psychology.

The authors of personality tests, or psychometricians, as they later became known, aimed at identifying personality traits with a view to measuring them. However, serious problems were encountered in this regard. According to Cattell (1965: 55), there were just as many traits and interpretations of traits as there were psychologists. He refers to the finding by Allport and his co-worker, Odbert, of more than 4 000 dictionary definitions of personality traits. However, Cattell pointed out that personality comprises natural unit structures and that these structures, rather than the endless names found in dictionaries, should be the point of focus. The development of correlation and factor analysis enabled psychologists to bring order to and to find structures in the confusion of concepts surrounding personality traits. In this way the scientific measurement of personality was eventually placed on the road to meaningful development.

Although personality is defined as including the entirety of human behaviour, a distinction appears to have arisen in the course of the development of personality psychology between the fields of study of the cognitive (or intellectual) and non-cognitive (or non-intellectual) aspects of personality. The study of the non-cognitive aspects became known as personality studies. Most personality tests exclude the measurement of general intelligence and aptitude (intellectual aspects) and concentrate on the dynamic and structural aspects of personality, such as interpersonal relationships, motivation, interest, attitudes and emotions. In time these aspects became synonymous with *personality*. However, for a full personality evaluation information on both cognitive and non-cognitive personality

traits should be integrated. The psychologist should know what an individual can do with his intellectual ability, for example whether he will leave it unused owing to lack of motivation, or, conversely, whether he just does not have the intelligence to realize his objectives, despite strong motivation.

Personality tests as measuring instruments for the non-cognitive aspects of personality can be further divided into two categories, measurement by questionnaire techniques and measurement by projective techniques.

A scientifically developed questionnaire consists of a number of questions or items that are tested and selected in such a way that a high degree of reliability, factorial purity and at least construct validity are obtained. However, the actual value of the questionnaire depends to a great extent on the *bona fides* of the respondent or testee. The testee may, for example, realize the aim of a questionnaire and deliberately formulate answers to meet this aim. A projection test, on the other hand, uses unstructured, ambiguous or multivalued stimulus material, such as ink blots (Rorschach) or pictures (TAT) that depict human situations in such a way that various interpretations of the situation are possible. Because the respondent does not know the aim of the stimulus, she *projects* her own meanings into the stimulus. In this way she reveals something of her conscious and subconscious fantasies, feelings, desires, needs, values, motives, etc. The principle on which the projective technique is based is that everything we do bears the stamp of our personality to a greater or lesser degree. One's personality therefore also influences one's perceptions of things. The value of the projection test, as compared with the questionnaire, lies in the fact that it is less vulnerable to deliberate manipulations by the respondent.

The advantage of a questionnaire, on the other hand, lies in the fact that its scoring and interpretation are generally more reliable and objective than in the case of a projection test. For example, a questionnaire can be scored with a stencil, while the scoring or evaluation of projection test responses is generally highly dependent on the experience, insight and skill of the psychologist.

Depending on the purpose for which the test was developed, a personality test measures certain constructs that have usually been identified on a theoretical basis. Through the specific formulation of questions, constructs such as introversion-extraversion and dominance-subjection can be *incorporated* into a personality questionnaire. In the same way projection tests can measure different constructs through the specific design of the stimulus material. For example, a picture of a man and a woman will elicit responses describing a man-woman relationship from most respondents.

The cards or pictures of a projection test developed for clinical purposes will include constructs such as attitude towards authority, recognition and channelling of aggression, sense of responsibility and leadership. Projection tests developed for use on children often use animal figures such as bears, rabbits and cats to measure personality traits such as parent dependence, fear of or liking for school and sociability. Thorndike *et al.* (1991: 408) pointed out that measures of personality have been developed for two somewhat different purposes. On the one hand some inventories, such as the Sixteen Personality Factor (16PF) Questionnaire, try to describe the normal-functioning person and to give guidance in dealing with minor problems of adjustment. On the other hand, such instruments as the Minnesota Multiphasic Personality Inventory (MMPI) focus on those with more severe problems and seek to diagnose serious mental disorders.

In conclusion it should be said that psychology cannot lay claim to the same degree of measurement accuracy as that attained in the physical and biological sciences. The human psyche is much too complex for that. As far as personality and motivation tests are concerned, Cattell (1983: 254) declares:

> It has taken about half a century to reach the same level of clarity in regard to personality structure as was achieved by Spearman and Thurstone in the first 30 years of this century in the field of abilities. What we recognize now in the personality field is that no matter whether one approaches by ratings of behavior in everyday life or by questionnaires, or by situational performance test of personality, one

> arrives at roughly some 20 primary factors and some 8 or 9 secondary (second order) factors. It is still not always clear what the origins of these separate structures are. Among the primaries we recognize proof of Freud's notion of an ego structure and a super ego structure, of Bleuler's conception of a schizothyme and cyclothyme temperament dimension, of Jung's notions of extroversion and introversion, as well as some half dozen factors which could not be perceived at the clinical level but required the microscope of multivariate, factor analytic methods. The same structures have been shown to exist at different age levels, developing through childhood, and also in different cultures, in that structure of the *16 Personality Factor Questionnaire*, the *Clinical Analysis Questionnaire*, and the *High School Personality Questionnaire* all seem to be much the same in Anglo-Saxon countries, in France, Italy, Germany, Japan, etc. We can thus conclude that we are dealing with essentially universal dimensions of human nature.

Analyses based on psychological tests should be regarded as supplementary information with high validity and reliability rather than as the *true* profile of a person's abilities or personality problems. It should also be kept in mind that validity and reliability are always calculated for *groups*. The psychologist must therefore act very carefully and responsibly when using tests to give advice or take decisions regarding a *specific individual*.

4. APPROACHES TO THE ASSESSMENT OF COGNITIVE DEVELOPMENT

Cognitive development focuses on the individual's ability to construct or understand reality. Terms such as cognition, thinking or intelligence are all aimed at defining an individual's problem-solving ability. In this process the impact of social context on individual achievement cannot be disregarded. There are three main approaches, which differ from each other in terms of item format, item selection, scoring criteria and clinical interpretation, to the assessment of cognitive development. These approaches are the psychometric, Piagetian and Soviet-based assessment techniques (Hoy & Gregg 1994).

The major categories of instruments that are most fully developed as practical assessment devices are, according to Daniel (1997:1038), psychometric-ability measures, neuro-psychologically based tests, and dynamic assessments.

4.1 PSYCHOMETRIC APPROACH

The factor analytic basis of the psychometric approach was outlined in paragraphs 3.1 and 3.2. The psychometric approach to the assessment of cognitive development involves the following (Hoy & Gregg 1994: 137-142):

- standardized procedures in test administration
- standard presentation of test items
- greater emphasis on the subject's product scores than on strategies used to obtain the answer to a problem
- little feedback from the examiner or tester
- items for the test are selected on strength of statistical criteria and correlation with a total test score is crucial to item selection
- for tests derived from factor theory, items must correlate either with g or with predominant item clusters (a item cluster is labelled as measuring a given trait, e.g. verbal, space, memory)
- interpretation of test scores is primarily quantitative

- the total of the subject's scores is converted to a standard score (IQ/percentile/ stanine, etc.) based on the distribution of scores obtained by the standardization sample

Some of the tasks involved in intelligence testing are:

- defining the meaning of words (vocabulary)
- understanding paragraphs (comprehension)
- organizing stimuli to show a progressive relationship (sequencing)
- completing analogies ("A is to B as C is to?")
- abstract reasoning (e.g. absurd verbal statements)
- memorizing stimuli (memory)

The psychometric approach has significant limitations, especially for the severely cognitively disabled, physically disabled and culturally different students. A central concern, however, that lies

> at the heart of many criticisms of the psychometric approach is that standardized IQ tests are used to allocate the limited resources of our society... Intelligence test results are used to provide rewards or privileges, such as special classes for the gifted, admission to college or advanced study, and jobs. Those who do not qualify for these programs may readily direct their anger at the tests because they see the tests as denying them opportunities (Hoy & Gregg 1994: 140-141).

4.2 PIAGETIAN APPROACH

According to Piaget (Hoy & Gregg 1994: 134-135), intelligence does not develop linearly but is constructed by successive stages of development - each stage reconstructs the previous one at a higher level of abstraction. Each stage is characterized by a certain view of the world and the child's relationship to it. The first stage is the *sensorimotor* stage (approximately from birth to 24 months). During this stage the child becomes aware of objects and the difference between objects and self. The next stage is the *preoperational* (approximately from 24

months to 7 years), during which the child begins to represent things to himself and to understand cause and effect relationships; reasoning is limited to those things the child can see and handle. In the following stage, *concrete operations* (approximately from 7 to 11 years), the child develops the ability to think independently of perceptions or how objects look. During the final stage, *formal operations*, the previous stage comes to fruition. The young individual can reason logically, form hypotheses, explore consequences and use more abstract reasoning (such as inference and figurative language).

The Piagetian approach, in contrast to the psychometric approach, does not focus on individual differences. Other features of the Piagetian approach (Hoy & Gregg 1994: 144) are:

- test administration makes use of a structured interview with the testee that focuses on the task stimuli
- qualitative analysis of the person's reasoning is used rather than quantitative
- attention is given to wrong answers as well as to right ones
- specific mental operations are emphasized rather than general intelligence (therefore most tests are constructed according to age levels and differ in this regard from the psychometric approach that uses subtests with different levels consisting of different types of items)
- reasoning abilities are investigated, and in scoring the emphasis is thus on the child's response, without any concern for speed (again in sharp contrast to the psychometric approach)
- researchers tried to build reliability into the assessment tasks by means of structured response formats and scores for mental and chronological age as well as quality of error.

Interpretation and administration of Piagetian tests cannot be done without mastery of Piagetian theory. According to Hoy and Gregg (1994: 144-145),

> It is assumed that the items pertain to actual operational mechanisms that govern behavior; this contrasts starkly with the psychometric

approach that deals only in vague, global capacities rather than structures of the intellect. As Piaget (1952) wrote, "it is indisputable that (traditional) tests of mental age have, on the whole, lived up to what was expected of them: a rapid and convenient estimate of an individual's general level. But it is no less obvious that they simply measure a 'yield' without reaching constructive operation themselves". In accordance with this assumption of direct measurement, individuals are classified as sensorimotor, preoperational, concrete operational, or formal operational based on test results.

The following are samples of Piagetian diagnostic tasks:

- Conservation of Number - measures the understanding that a specific arrangement of a row of objects does not affect the number of objects.

- Conservation of Continuous Quantity: Solids - measures the understanding that the quantity of a solid is not changed by variations in the shape of that solid.

- Conservation of Weight - measures the understanding that variations in the shape of an object have no affect on the weight of that object.

- Seriation: Size - measures the understanding that objects can be arranged in a certain order according to their size.

The Piagetian approach and techniques developed within this approach have been fruitfully applied in assessing preschool children. Valuable diagnostic information can be obtained by means of these techniques. The Piagetian approach has, however, certain limitations. According to Hoy and Gregg (1994: 145),

a problem with this assessment approach is that the tasks are often

divorced from real-world activities and provide very little observation of social interaction skills. Piagetian theory has also been criticized for its assumption that learning develops in a hierarchical progression, particularly whether children do develop cognitively only after the sensorimotor period ends.

Recently there has been renewed interest in the Piagetian approach as a means of developing assessment and intervention plans for adolescents and adults with special needs.

4.3 SOVIET APPROACH

Soviet psychology was responsible for a new concept in the field of intelligence, *the proximal zone of development*. The proximal zone is defined as "the distance between the actual developmental level as determined by independent problem solving, and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky in Hoy & Gregg 1994: 134). The work of Vygotsky, a Russian psychologist, and Luria, a Russian neurologist, provided the incentive for much of the current research and development regarding information-processing models of intelligence.

The content of Soviet assessment batteries used in diagnosis is equivalent to that of American psychometric instruments. The methods and emphasis of testing, however, diverge sharply. In Soviet assessment, a distinction is made between actual level of development, as indicated by the scores on a psychometric test, and potential level of development, as determined by the width of the proximal zone. Two individuals with identical test scores are therefore not considered to have equal ability because their proximal zones may differ. Hoy and Gregg (1994: 147) explain the difference between the two scores in the following way:

> [A] child with significant cognitive limitations cannot put a puzzle together. Another child from a culturally different background might

also have difficulty putting the puzzle together. Each of these children would score low on a standardized assessment measuring perceptual organization and reasoning. The Soviet approach, however, states that this standardized score is only the low end of the child's potential. If provided guided instruction or cues from the teacher, the student with a different cultural background would more than likely require fewer hints on how to develop strategies to complete the task than would the child with severe cognitive limitations. The amount and type of guided instruction should be part of a cognitive assessment.

The psychologist must therefore look at the testee's range from independent learning to mediated help, i.e. hints received, during a problem-solving task. It is this range that Vygotsky called the *zone of proximal development*. The philosophy behind this approach is that guided learning provides a more accurate measure of "true" potential than do the static tasks presented by psychometric measures.

The Soviet approach to testing proceeds as follows:

- The individual solves independently tasks similar to those found in American IQ tests.

- If difficulties are encountered, the tester (mediator) give progressively more cues and ascertains how many cues (bits of information) the testee needs to successfully answer the question or solve the problem.

- When the testee has completed a task successfully, another form of the original task is presented to him or her in order to observe transfer to a novel situation.

- The width of the proximal zone is obtained by comparing the number of cues needed to solve the second problem with the number of cues needed to solve the first one.

- Aspects taken into account in the scoring process are the original level of development, the number of cues needed to solve a problem and the degree of transfer.

The Soviet approach outlined above serves as the theoretical underpinning of dynamic assessment discussed in Par. 4.5. In contrast to traditional diagnostic procedures which have been static, leading only to diagnostic labels and placement decisions, the Soviet approach is much more dynamic.

4.4 NEUROPSYCHOLOGICALLY (BIOLOGICAL) BASED INSTRUMENTS

Luria's (1973) theory of the organization of abilities is, according to Sternberg (1997b:1134), at the heart of current attempts to build cognitive ability tests based on neuropsychological theory. Luria's model posits three functional levels, each associated with a region of the brain: at the lowest level are arousal and attention; at the next level, information is encoded and processed in either a simultaneous or successive manner; at the highest level, planning and monitoring functions take place (Daniel 1997:1040).

These four main cognitive functions, namely planning, attention, simultaneous processing and sequential processing, form the basis of the PASS theory of Das *et al.* (1994). The test developed by Das and Naglieri, based on the PASS theory, is known as the Cognitive Assessment System (CAS). According to Naglieri (1997:248) the "PASS theory, and the CAS, are the result of the synthesis of neuropsychology, cognitive psychology, and psychometrics with the emphasis on a theory-based theory of human cognitive functioning that includes a broad spectrum of measurement". Another test also based on the PASS theory, is the Kaufman Assessment Battery for Children (K-ABC) (Kaufman & Kaufman 1983).

These two tests differ markedly in outlook, development, and interpretation from those based on the psychometric-ability tradition (Daniel 1997:1039). The main difference is that psychometric-ability tests are constructed around models that

have grown gradually as empirical evidence has accumulated, while the content of tests based on Luria's neuropsychological model is more theoretically driven. The developers of CAS also believe that the problem of diagnostic differentiation has been poorly addressed by traditional IQ tests (Naglieri 1997:263). These authors maintain that their tests represent an improvement in this regard, enabling the user to differentiate between reading disabled, mentally retarded, attention deficit, delinquent and normal children. On the strength of the finding that different samples of children with different diagnoses had different PASS profiles from matched control groups, Naglieri (1997:263) declared that these data "illustrated the sensitivity of PASS and the advantages it may provide because some of the cognitive processes we measure are not assessed by traditional IQ tests". Furthermore, "PASS may offer a viable method for solving the problem of differential diagnosis and may provide the opportunity to consider a theoretical view of what the disabilities may be".

Daniel (1997:1040) responded inconclusively to the question whether the tests based on the PASS theory (CAS and K-ABC) generated scores that provide new information or a reconfiguration of psychometric abilities. The constructs of the PASS and the psychometric systems overlap to some extent as is evidenced by the fact that subtests from the two systems correlate. This does not, however, mean that subtests designed according to the neuropsychological model cannot measure a different system of processes as well. Daniel (1997:1040) concluded that as in the case of "psychometrically oriented intelligence tests, research clarifying the constructs represented by scores on the neuropsychologically based tests would be worthwhile". And further "... as with the psychometric-ability batteries, construct validation is needed".

4.5 DYNAMIC ASSESSMENT

Most conventional tests of abilities, i.e. tests that are based on psychometric theory, are static and no feedback is given to the testee.

Dynamic tests, on the other hand, offer a new option for measuring abilities

(Sternberg 1997b). According to Lidz (1997:281), dynamic assessment "refers to approaches to the development of decision-specific information that most characteristically involve interaction between the examiner and the examinee, focus on learner metacognitive processes and responsiveness to intervention, and follow a pretest-intervene-posttest administration format".

The work of Vygotsky (1978) together with that of Feuerstein (1980) and others, has provided "the theoretical and empirical base for restructuring the purpose and means of assessment in special education" (Hoy & Gregg 1994:148). Dynamic assessment procedures attempt to provide the following types of information (Daniel 1997:1040): (i) more valid measures of the abilities that are measured by static tests; (ii) measures of various abilities, particularly learning ability or modifiability; (iii) understanding of the cognitive processes the student uses or fails to use; (iv) clues about the instructional methods that are most effective for the student.

Dynamic assessment developed, according to Lidz (1997:281), "both as a reaction to dissatisfaction with existing procedures as well as a positive attempt to design a model that is theory-based, provides a meaningful description of cognitive functioning, and links assessment with instruction". The rationale behind dynamic assessment is that if you wish to understand how a student learns, it is best to engage the student in the learning process. This approach is related to the view, attributed by Vygotsky to Marx, that a phenomenon or process can best be understood when one tries to change it (Lidz 1997:281). Dynamic assessment presents a situation in which the student engages in the learning process while the examiner attempts to facilitate his/her cognitive competence. The interaction between the examiner and the student serves not so much to sample typical functioning as to optimize such functioning. In the case of intelligence, for example, it is not the assessment thereof that is important but the observation of the application of "intelligence" or intelligent functioning within the learning situation. Dynamic assessment yields information on how the testee profits from assistance, the testee's speed of learning and the testee's generalization abilities. It also provides the clinician with a wealth of information on intra-individual

functioning in different situations as well as on certain inter-individual comparisons.

In contrast to the neuropsychological approach to intelligence testing which presents a conceptualization of abilities that is an alternative to the psychometric model, the dynamic assessment approach is less concerned with the structure of abilities but focuses more on a different aspect of intelligent behaviour, namely the ability to learn (Daniel 1997:1040). Dynamic assessment cannot be viewed as another instance of psychometric assessment; it involves a paradigm shift, both in the conceptualization of cognitive functioning itself and in the approach to assessment. According to Lidz (1997:292), dynamic assessment is a genuinely different approach, not only with a different methodology but also with different assumptions. Although the model assumes that learning is a process of change and the result of interaction, no assumptions are made about how much can be learned; nor can outcomes for individuals be predicted with confidence from current or previous performance.

Dynamic assessment begins where traditional psychometric assessment ends; the results of most traditional procedures represent the starting point (i.e. the pretest) of dynamic assessment (Lidz 1997:282). A dynamic assessment test (or learning potential test) has the psychometric properties of a conventional test but differs from it with regard to its administration procedure as a training phase is incorporated. This phase is usually preceded by a pretest and followed by a posttest. The improvement in performance from pretest to posttest indicates the testee's learning potential. This score difference between pretest and posttest not only reflects the testee's ability to profit from guided feedback but is also an indication of the difference between the testee's latent capacity and his/her observed ability. Vygotsky (1978) referred to this as the zone of proximal development (ZPD). The primary guiding principle of dynamic assessment, according to Lidz (1997:282), is Vygotsky's view that ZPD is an integral component of assessment, together with the zone of actual development. The zone of actual development describes the testee's independent level of performance, whereas the zone of proximal development describes what the testee is able to achieve with the help of an experienced collaborator. This collaborator

may be any person (e.g. a teacher, a parent, a peer, or a sibling) with more experience in the particular domain.

Supporters of the dynamic assessment procedure believe that it is a viable way of approaching culturally different and disadvantaged populations, and that it is especially suitable for deprived children, children with learning difficulties and children from ethnic minorities (Hamers & Resing 1993:27). They argue that conventional ability testing can result in an underestimation of these children's real intellectual potential; the training phase in the test is a means of offering children an optimal chance of achieving a fair test result.

Dynamic assessment procedures can be divided into two groups according to the way the tests are administered (Daniel 1997:1041). One group uses clinical, non-standardized intervention by the tester to reveal the cognitive processes in which the testee is weak, to identify effective intervention methods and to improve the testee's cognitive processes. In the clinical versions of dynamic assessment not much attention is given to the psychometric properties of the test scores (Daniel 1997:1041). The best-known example of this type of dynamic assessment is Feuerstein's (Feuerstein, Rand & Hoffman 1979) learning potential assessment device (LPAD). The instruments in the LPAD "serve to provide mediated learning experiences that create a zone of proximal development and allow observation of the student's facilitated functioning" (Lidz 1997:283).

The other group, according to Daniel (1997:1041), consists of techniques that provide standard rather than clinical interventions, that use objective measures of the number and type of prompts (hints) required, and that give an indication of the amount of growth following intervention. Although several researchers are involved in this particular branch of dynamic assessment (e.g. Campione & Brown; Embretson; Guthke & Stein — see references in Daniel 1997:1041), the only normed instrument of this kind published in the United States is the Swanson Cognitive Processing Test (Daniel 1997; Sternberg 1997b).

What then are the limitations of dynamic assessment? Suzuki and Valencia

(1997:1111) and Hamers and Resing (1993:37) point out that empirical research has been done only during the past decades and that attempts are still in their infancy. According to Lidz (1997:286), dynamic assessment is a "psychometrician's nightmare" because traditional notions of reliability (especially test-retest) are not automatically relevant — appropriate psychometrics may have to be developed. Reschly (1997:449) believes that in respect of a number of important issues — the fact that all learners appear to have more potential than is demonstrated in actual performance, the accurate classification of cognitive structures, the estimation of the modifiability of these structures, and the time or effort required to produce the modifications — the evidence to date is not convincing. "Much work is still needed regarding the technical adequacy of dynamic assessment" (Reschly 1997:449).

Dynamic assessment, together with procedures to investigate problem-solving and thinking skills, has led clinicians to re-evaluate diagnostic models (Hoy & Gregg 1994). Hoy and Gregg are, however, of the opinion that dynamic assessment should be used in conjunction with standardized, criterion-based and curriculum-based testing. Although dynamic assessment has many advantages, it is more suited to the clinical situation, i.e. to the individual case, than to the group-testing situation. In other words, traditional group tests such as multiple aptitude test batteries cannot easily be replaced by dynamic assessment techniques.

4.6 CONCLUSION

Attention was given here to the three main approaches to the assessment of cognitive development, namely the psychometric, Piagetian and Soviet-based assessment techniques. If one takes a closer look at the assessment of intelligence *per se*, and especially the theory behind the assessment techniques, three important categories of intelligence tests are apparent. These include psychometric-ability instruments, neuropsychologically based instruments and dynamic assessments — the latter two represent newer developments in their field of testing and were therefore discussed in some detail. Alternative conceptions of intelligence — not discussed in this document — can also be separated from

conventional kinds of psychometrically measured intelligence. These include emotional intelligence, practical intelligence and social intelligence (see Sternberg 1997a).

Twenty years ago (1978), Carroll (cited in Daniel 1997:1038) declared that "the present scene in intelligence testing is essentially one of stagnation, with much talk but little progress". Since then, some progress has been made with the emergence of the neuropsychological and dynamic assessment approaches. The relationship between the three assessment models (psychometric, neuropsychological and dynamic) is, in short, the following (Daniel 1997:1040): The neuropsychological approach to ability testing offers a conceptualization of abilities that is an alternative to the psychometric model; the dynamic assessment approach, in turn, is less concerned with the structure of abilities than the psychometric approach and is more involved with a different aspect of intelligent behaviour, namely the ability to learn. Dynamic assessment, like neuropsychological assessment, focuses on cognitive processes and emphasizes the teachability of those processes.

In conclusion, 80 years of research indicates that general intelligence — as assessed by means of the psychometric-ability model — is the best predictor of performance in training and performance later in the job (Ceci & Williams 1997:1051). This long period of empirical findings gives the psychometric-ability model a "type of robustness that more theoretically driven models do not enjoy to the same degree" (Daniel 1997:1043). Although the psychometric-ability model will probably continue to enjoy prominence, Daniel (1997:1043) points out that it is facing increasing pressure to show the practical application and benefits of abilities in educational, occupational and clinical fields. At the same time the newer tests, based on alternative models, are not necessarily in a stronger position. For these tests to replace psychometric-ability tests, their champions "need to do more than point out the weaknesses of such instruments; they also must demonstrate that the new tests provide one or several practical benefits that are superior to what psychometrically based tests can offer" (Daniel 1997:1043).

5. PSYCHOLOGICAL TESTING: CRITICISMS, ISSUES AND CONTROVERSIES

In a certain sense, psychological testing is the victim of its own success. In this regard Zeidner and Most (1992) have pointed out that, despite the enormous advances made in psychological tests since the beginning of the 20th century, their phenomenal growth in number, variety, and functions and increased usage in decision making have brought them under scrutiny and attack. There may be several reasons for this, one being that the indiscriminate use of tests has inevitably led to misapplication and the general misuse of test results and another being that certain misconceptions of what can possibly be achieved with tests may have led to inappropriate or unjust criticisms or actions against tests. Be that as it may, the recent controversies surrounding ability, personality and vocational tests are reminiscent of debates from the beginnings of modern testing, with the same misconceptions, the same value conflicts and the same arguments continually resurfacing (Cronbach 1990; Jensen 1980).

5.1 CRITICISMS

Some of the recurring criticisms of psychological tests are:

(1) Psychological testing is conducted too frequently and often without sufficient justification.

> It is claimed that students are required to take a bewildering array of aptitude, achievement, occupational, and personality tests throughout the school years, when time might be spent more usefully on other activities. Furthermore, testing is often carried out without any clear purpose in mind or where better measures of the criteria of interest are easily available. A case in point would be a school psychologist who administers scholastic aptitude tests to eighth-grade students to predict their academic performance in ninth grade, despite having their cumulative grade point averages through grade eight to refer to (Zeidner & Most 1992: 40-41).

(2) Psychological tests are often claimed to be an unwarranted invasion of privacy.

In applying for a job, examinees have sometimes been required to divulge personal information that has little to do with success on the job. At one end of the spectrum are tests of job knowledge, skill or ability to which no one is likely to object when that attribute is clearly linked to the particular job. At the other end of the spectrum are self-descriptive instruments that lead to inferences about emotional stability, honesty, hostile feelings, anxiety, etc.; testees are led to give this intimate information without knowing how it will be used. Certain ethical concerns have also been raised concerning psychological tests used in basic research. These involve invasion of privacy, deceptive ploys and causing psychological harm through aversive test instructions or test content. In an attempt to give direction in this regard, the courts in the United States are to an increasing extent taking a role in deciding what information is allowable. Recent court decisions have required a demonstration of the validity or relevance of test scores or personality profiles to job performance before such instruments can be used for employee selection, and the decisions have also affected the type of information that may be acquired by limiting invasion of privacy. Another aim of the court decisions is to advance the causes of affirmative action and antidiscrimination (Thorndike *et al.* 1991; Zeidner & Most 1992).

(3) Psychological tests, particularly maximal performance tests, tend to evoke anxiety.

In general, it appears that small amounts of test anxiety may have a facilitating effect on test scores while higher levels may have a detrimental effect on performance. According to Zeidner and Most (1992: 41), support was found for the notion that in situations where the individual will be judged, highly test-anxious individuals direct their attention away from the task at hand to self-related cognitions which hamper their performance. Although it cannot be denied that a chronically high anxiety level exerts a

detrimental effect on school learning and intellectual development, the important question is: "To what extent does test anxiety make the individual's test performance unrepresentative of his or her customary performance level in nontest situations?" (Anastasi 1990: 41). According to this author, research shows that students who score high on a test anxiety scale obtain lower grade-point averages and tend to have poorer study habits than do those who score low in test anxiety. In this particular case, test anxiety is not caused by the test as such but is part of the person's emotional make-up. With regard to the nature of test anxiety, a distinction should be made between emotionality and worry (Anastasi 1990). Emotionality has to do with feelings and physiological reactions, while worry involves negative self-oriented thoughts such as being afraid of doing poorly and concern about the consequences of failure. These thoughts tend to disrupt performance.

On the whole, test anxiety has not figured prominently among the variables hypothesized to account for cultural or racial group differences in test scores (Jensen 1980: 615).

(4) Tests are mainly used to serve the decision-making needs of the user institution and not the needs of the test taker. The person-centred functions of testing are often treated as byproducts or afterthoughts (Messick 1982).

(5) There is strong evidence that tests create self-fulfilling prophecies, that is, can influence and precondition teacher expectations of children.

(6) Teachers do not understand the meaning of the results obtained from psychological tests.

(7) An important observation that can be made about the controversy over testing is that the underlying dissatisfaction is often not with the tests themselves, but with the social functions tests are playing. For example, there is debate over special education placement. What is really under

attack is the overrepresentation of minorities in programmes perceived to carry a negative label and to offer little in the way of improved educational services. Selection in higher education would evoke uneasiness in a society used to viewing education as a right and as a basis for personal and social mobility (Resnick & Resnick 1982: 86).

(8) Tests of abilities do not necessarily give a true picture of a given individual. Consequently, they have been criticized for serving the needs of the organization more than those of the individual. Tests have also been criticized for perpetuating cultural, gender and socio-economic bias.

(9) There is a lack of a satisfactory definition of what is tested.

This is to a large extent true, but in many instances nevertheless an exaggeration. The fact that psychologists cannot agree upon an exact definition of a particular construct or abstraction does not mean they have different things in mind when referring to that construct. Intelligence is a case in point. Although there are more than 24 different theories of intelligence, there is a remarkable agreement among professionals as to what intelligence is. Snyderman and Rothman (Li 1996: 6-7) surveyed 1 020 social scientists and educators on many topics dealing with the nature of intelligence. Surprisingly strong consensus was found among scholars on the elements of intelligence. As Snyderman and Rothman put it (Li 1996: 7),

> Accompanying the disagreement about the scope of the definition of intelligence is very strong agreement at its core. It can reasonably be concluded that when different psychologists and educators use the term 'intelligence' they are *basically* referring to the same concept, having to do with the capacity to learn and with more complex cognitive tasks like abstract reasoning and problem solving, and that they would generally exclude purely motivational and sensory abilities from this definition.

(10) Tests are based on the cultural experience and operate through the language of the dominant cultural group.

Many allegations have been made regarding the inappropriateness of using tests, and in particular intelligence tests, with bilingual individuals (Hoy & Gregg 1994). It can hardly be denied that a student's race, social class and primary language all influence performance in an intelligence test. A person with limited English proficiency who is being evaluated in a situation in which English is the primary language is at risk of inaccurate assessment of his or her true ability. However, a distinction should be made, according to Anastasi (1990: 64), between cultural factors that affect both test and criterion behaviour and those whose influence is restricted to the test (it is the latter, called test-related factors, that reduce test validity). The specific test content may influence test scores in ways that are unrelated to the ability the test is designed to measure. Anastasi (1990: 65) illustrates this in the following way:

> In a test of arithmetic reasoning, for example, the use of names or pictures of objects unfamiliar in a particular cultural milieu would represent a test-restricted handicap. Ability to carry out quantitative thinking does not depend upon familiarity with such objects. On the other hand, if the development of arithmetic ability itself is more strongly fostered in one culture than in another, scores on an arithmetic test should not eliminate or conceal such a difference.

In the last sentence quoted here, Anastasi makes a very important point that is often overlooked by those who criticise tests, that test score differences between individuals or groups are not in themselves evidence that the particular measuring instrument is biased: the differences could be real.

(11) Standardized tests are biased in content, procedure and use; an unbiased test is virtually impossible, almost a contradiction in terms.

The issue of bias in testing is certainly the most hotly debated topic regarding the development and use of psychological and educational tests over the last 25 years (Thorndike *et al.* 1991: 457). According to Zeidner and Most (1992: 42), bias "has become a key villain in the drama surrounding the use of psychological tests." The allegation is made that, because of bias in standardized tests regarding content, procedure and use, these tests have questionable validity not only for assessing the intellectual abilities of minority groups but also for predicting these groups' future performance on a criterion. What gave rise to this allegation (in the USA) is the repeatedly observed group differences in favour of majority group testees in intelligence, aptitude and achievement test scores. A number of situational variables in standardized test administration and content thought to be detrimental to the test performance of minority groups, such as test attitudes, examiner-examinee rapport, the race of the tester, testing time limits, motivation and anxiety, were investigated. Tests were also scrutinized for bias by means of judgmental and statistical techniques. The results of all these investigations show that the

> currently most widely used standardized tests of mental ability - IQ, scholastic aptitude, and achievement tests - are, by and large, *not* biased against any of the native-born English-speaking minority groups on which the amount of research evidence is sufficient for an objective determination of bias (Jensen 1980: ix).

The topic of test bias will be explored in more detail in paragraph 5.2.

(12) Other points raised by critics are (Zeidner & Most 1992: 43):

- Standardized multiple-choice test items are often ambiguous and have more than one correct or justifiable answer.

- Tests reward students with only partial knowledge, penalize bright

and creative testees and are insensitive to atypical but defensible responses.

- Tests measure only limited and superficial aspects of knowledge or behaviour and are unable to measure truly important characteristics. Test users therefore make decisions on the basis of relatively unimportant and superficial information.

- Test usage leads to undesirable attitudes, since many believe that psychological measurements are infallible and test performance has to do with something innate that cannot be modified. Thus, teachers and parents regard IQ or aptitude scores as accurate, unmodifiable measures and treat children according to tested expectation levels, disregarding other information.

To conclude, it is evident from the above that psychological tests are not without problems and hazards and many of the criticisms are certainly warranted. On the other hand, one should not lose sight of the fact that tests also serve important functions in a wide variety of situations.

5.2 TEST BIAS

Since the earliest measurement of the intellectual ability of human beings, it has been evident that tests can be class or culture linked. As early as 1905, Binet and Simon noticed that on their new test of "intelligence", Parisian children of high social status scored better than children of the lower or working class. Similar differences were found in Belgium, Germany and the United States (Owston 1984: 47). Before publishing the second revision of the Binet-Simon Scale six years later, Binet removed test items that he thought contributed to the differences between the classes. This stratagem was, however, unsuccessful and the apparent bias of the test against the lower classes still persisted in the revised version.

According to Reynolds and Brown (1984), the question of bias in intelligence tests arose mainly as a result of the nature of psychological processes and the measurement of such processes. Psychological processes are not directly observable or measurable and consequently have to be deduced on the basis of behaviour. In psychology there is consensus on very few of these deductions or hypothetical constructs. It is therefore understandable why intelligence or intellectual ability - surely one of the most complex processes in psychology - has attracted the interest of experts as well as laypeople. It is against this broad background that the criticism (including that of bias) levelled against psychometric tests of intelligence, aptitude and so on should be seen. The accusation of bias is heard particularly from minority groups in the United States and elsewhere who maintain that the particular tests are more suited to the group with the largest share in the standardization sample. Whether such tests are actually biased and prejudicial to minority groups is one of the questions that has to be answered empirically.

In order to limit the effect that cultural circumstances may exert on test achievements, Cattell in 1940 proposed a "culture-free" intelligence test. In the United States the first systematic investigation into cultural bias in psychometric tests was undertaken in 1945 by Allison Davies, a sociologist, and Kenneth Eells, a psychologist. Culture bias in test items was regarded by these investigators as only one of several reasons - including inherited characteristics, development factors, motivational factors, working habits and skill in writing tests - for the differences in average IQ scores between different cultural groups.

During the 1960s the issue became more pressing in the United States with concern being expressed about the fact that the test achievements of blacks and other minority groups (the "culturally disadvantaged") were on average poorer than those of white Americans. These differences were apparent in a wide variety of tests, for example intelligence, scholastic aptitude and achievement tests. Because they indicated differences, the tests were said by (in particular) sociologists, anthropologists, educationists and other critics not actually acquainted with the field of psychometrics to be culturally biased against blacks. The upshot was the

formation of pressure groups which strove for the total abolition of psychological tests. However, responsible people and institutions realized that the fair use and possible bias of tests should be carefully looked at instead of their simply being abolished.

During the 1970s psychometricians also began to examine concepts such as test bias and test fairness more systematically. Not only did this bring greater clarity in respect of definitions and terminology but it also stimulated considerable research.

What is test bias?

"It is typically defined as the systematic error of some true value (e.g., test scores) of individuals that are connected to group membership ... such membership would be along lines of race and ethnicity. Bias in the context of racial or ethnic membership is typically referred to as *cultural bias* (Zeidner & Most 1992: 403). According to Anastasi (1990: 194), the term "bias" refers to constant error as opposed to chance error; it is in this sense that we speak of a biased sample, in contrast to a random sample.

In mathematical statistics the term *bias*, according to Jensen (1980), refers to a systematic underestimation or overestimation of a population parameter by a statistic based on a sample from that population; in psychometrics bias refers to systematic errors in the *construct validity* or the *predictive validity* of the test scores of the individuals associated with the group membership of those individuals.

These two aspects of test scores, namely construct validity and predictive validity, represent the two main areas of research on test bias. Superficially, the two areas are divergent and have very little in common regarding methods and techniques of investigation. However, they do not represent different concepts of bias.

Construct bias

Construct comparability is the most basic or fundamental question because it concerns the nature and essence of what is being measured: is the same construct or psychological dimension being measured in the various cultures? It may also be asked whether the particular construct occurs in the other culture.

Construct *bias* therefore means that the test measures something else in one group from what it measures in another group, while it is assumed that the same construct (for example, intelligence, mechanical insight or musical aptitude) is being measured.

The following indicates the *absence* of construct bias in an instrument:

- similar test reliabilities in the two cultural groups
- similar rank orders of item difficulty values
- similar item discrimination values
- similar factor structures

From a large number of investigations conducted in the United States, it appeared that some of the best-known ability tests, such as the Stanford-Binet and Lorge-Thorndike, reveal similar factor structures and test reliabilities for widely divergent groups (see for example Jensen 1980). Locally the same tendency is revealed in respect of, for example, the Junior Aptitude Tests (JAT) (Owen 1991). It therefore appears that there is a greater correspondence between the cognitive structures of the various cultural groups than is often thought.

However, the absence of construct bias does not exclude the possibility that the ability of a group may not be systematically underestimated by biased items, whether as a result of language or of other factors. When language as such is measured, the differences between the groups possibly indicate an actual difference in respect of the particular ability; differences in this case are therefore not necessarily a function of bias. An alternative explanation would, for example,

mean that all spelling tests are biased for poor spellers or that all arithmetic tests are biased for those who cannot add or subtract.

In conclusion, with regard to bias in the construct validity of South African ability tests, it can be expected that most tests measure the same constructs in different groups - which have more or less the same scholastic qualifications - and that bias will not occur often. It can also be expected that tests that contain language will be less reliable than tests that do not contain language. This point is extremely important for non-cognitive tests, such as personality questionnaires, which consist exclusively of language. Those who use these types of tests will have to consider the language proficiency of the testees carefully when interpreting the test scores.

Item bias

Score comparability can be investigated meaningfully only *after* construct comparability has been shown. What is involved here is whether a score t for one group is the same -in terms of the amount of the underlying characteristic or construct - as the score t for another cultural group.

Investigations into item bias are aimed at determining whether different cultural groups manifest different behavioural patterns in respect of test items. A typical statistical indication that a test item may not be suitable for a certain cultural group is whether the item is clearly too difficult (or too easy) for the group. According to one definition,

> An item or subscale of a test is considered to be biased in content when it is demonstrated to be relatively more difficult for members of one group than another when the general ability level of the two groups being compared is held constant and no reasonable theoretical rationale exists to explain group differences on the same item (or subscale) in question (Reynolds 1982: 188).

However, it should be emphasized that the poor achievement of a group does not

necessarily mean that the particular test or item is biased. A group of stutterers will, for example, always perform poorly in a fluency reading test. Bias is at issue only when persons or groups who have the *same* ability do *not* have the same chance to answer a particular item correctly.

Predictive bias

The third type of comparability, predictive comparability, can be evaluated only if a criterion is available. According to the Cleary (1968) definition, a test is biased if the criterion score, which is predicted with the help of the common regression line, is consistently too high or too low for members of the subgroup. Conversely, a test is unbiased if the regression lines of the groups are identical; in these circumstances group membership, such as race or sex, does not play a role.

The essential characteristics of predictive bias are that

- it is a type of invalidity that prejudices one group more than another group;
- group differences in test achievement are not reflected by corresponding differences in the behaviour domain that the test is meant to measure;
- it involves constant and systematic errors (e.g. attenuation as a result of the unreliability of the criterion), in contrast to errors that can be ascribed to coincidental or chance factors in the estimation of the criterion score (the constant or systematic errors are usually associated with group membership);
- it leads to unfair discrimination against the group whose criterion score is underpredicted - i.e. in practice the group does better in respect of the criterion than is predicted on the basis of the test scores.

In contrast to bias in the abovementioned connection, a test can be described as *unbiased* for a group if the deductions made on the basis of the test scores take

place with the smallest possible random error, or if the constant (systematic) errors do not occur in the deductions as a function of membership of a particular group (e.g. race or sex).

Regarding findings on bias in predictive validity, various researchers (e.g. Schmidt & Hunter 1981; Sackett & Wilk 1994) conclude that cognitive tests are equally valid for minority and majority groups (in the USA) and that the tests are *fair* towards minority groups in the sense that they do *not* underpredict the expected work achievement of these groups. On the contrary, a fairly general finding is that the work achievement of blacks in the USA is *over*predicted if the regression equation of whites is used, the cut-off of the regression line of whites being generally higher than that of blacks. Differences between groups in respect of test achievement are often also accompanied by differences in the criterion (work) achievement of the groups. The differences between the groups are therefore actual differences - they are not artificially "caused" by the tests.

As to bias in the predictive validity of South African tests, it can reasonably be accepted that American findings on similar tests will not be applicable to the same extent here. Because of the greater language differences that are found locally between the different groups, which *inter alia* calls into question the reliability of the tests, it can be expected that this form of bias occurs in the majority of tests. If a common regression equation, or the regression equation of the white testees, is used with the other population groups, the criterion achievement of the latter groups will most probably be overpredicted; in this respect bias is therefore actually in their favour.

During the 1970s and 1980s there were numerous empirical studies on test bias in the USA. Currently, however, there appears to be less interest in this kind of research. "The decline in test bias research", according to Suzuki and Valencia (1997:1109), "can be attributed, in part, to the consistent findings showing that prominent intelligence tests are not biased".

Fairness in testing

The central problem in the testing of different ethnic groups revolves around the question of unfair discrimination. Discrimination can be either fair or unfair: unfair discrimination occurs, for instance, when those with an equal chance to achieve success in a job do not have an equal chance to get the job. It is important to note that this factor (also known as *selection bias*) is not only a technical question but also involves *value judgments* in determining what "fairness" is.

Fairness therefore does not so much have a bearing on the characteristics of the test as on the *use* of the test. In this connection various writers, including Jensen (1980), stress the point that *test bias* and *test fairness* are two separate issues: unfairness is not seated in the test itself whereas bias is; a biased test can be used in a fair and an unfair manner and the same applies to an unbiased test. According to Jensen (1980), bias is a statistical judgment while fairness is a value judgment.

The concept of unfairness is based on a philosophical position concerning the way in which test scores, especially in education and personnel selection, should be used. Here it should be emphasized that the *social functions* of standardized tests were what landed testing in the United States in troubled waters.

A new development that has flowed from the concept of unfairness is the creation of fair selection models. Since a comprehensive literature on the subject already exists (e.g. Petersen 1980; Jensen 1980), some of the most important models will only be mentioned here:

- The regression model
- The quota model
- The equal risk model
- The constant ratio model
- The conditional probability model

It should be remembered that when the average test scores of two (or more) groups differ in respect of the same test and it is difficult to decide which candidates should be accepted, the use of selection models (provided there is an unbiased criterion) can to a certain extent help in treating a certain group more fairly.

In conclusion, as Jensen (1980) (see also Kline 1993: 165-166) has pointed out, there are three fallacies concerning the definition of test bias which must be summarily dismissed.

1. The egalitarian fallacy

This assumes that if any mean difference occurs between groups on a test, the test is necessarily biased. Although this argument is absurd and needs no further discussion, it nevertheless lies at the heart of much criticism of intelligence tests which reveal that blacks score lower than other groups. This is not saying that blacks are less intelligent but simply that it is a false assumption to use such data as evidence of test bias. Other data than mere group differences are necessary to make the point.

2. The culture-bound fallacy

This assumes that group differences on a test are due to the culture-bound nature of items. An intelligence test item, for example, based on what was common knowledge in one group but rare in another, would be assumed to be a source of bias. Items such as these are clearly biased, but the point stressed by Jensen is that it is impossible, without empirical evidence, to know which items are thus biased. It is necessary to determine on psychometric and statistical grounds whether an item is biased or not.

3. The standardization fallacy

It is often assumed that if a test is standardized on one group or population, it is

necessarily biased if it is used on another. Again this is not necessarily the case. Other evidence would be needed to decide whether the test was indeed biased in the new situation.

Although South African tests are generally reliable and valid, this applies mainly to the groups for which they were standardized. No test is *in itself* reliable or valid; however, it may be so in respect of a particular group. It is therefore up to the test user to determine empirically whether a test complies with the necessary psychometric requirements for the group he or she administers it to. In South Africa with its heterogeneous population we cannot permit a psychological test - especially when it is used for selection - to be used that reflects group differences that are irrelevant and invalid (which is essentially what test bias is, in other words, bias is the extent to which measured group differences are invalid). The fairness of a test and the just use of that test are in the final analysis the responsibility of the user.

The abovementioned problems that are experienced with cognitive tests in an intercultural situation also apply to a lesser or greater extent to personality measurements. The psychologist should never lose sight of the fact that all psychological instruments are based on one or other theory of human behaviour, all of which have their origins in the West.

6. CULTURE AND TESTING

6.1 INTRODUCTION

A particular culture stimulates a particular form of cognitive development, in other words, intellectual abilities are *culture bound* (cf. for example Lesser, Fifer & Clark 1965; Scarr 1981). The study of cultural differences in respect of intellectual abilities touches on a fundamental question in psychology: how can a valid psychological evaluation be made of persons from widely divergent groups? One reason why a definite answer has not yet been given to this question must be that the concept of, for example, intelligence is based on a Western technological culture: "It is not so much that tests are unfair to lower-status groups, as that lower-class environment is not conducive to the effective development of 'intelligence' as defined in our culture" (Lesser *et al.* 1965: 11).

The possible connection between culture and cognitive development is clear from the prerequisite that Ferguson (1954) sets for the origin of an ability: there must be the opportunity in a culture for the *over*learning of an activity. From this it follows that if a certain activity does not occur in a culture, there must likewise also be a lack of a certain ability.

Various writers point out that although the components of the cognitive system (memory, categories, associations, coding and decoding, semantic integration and verbal explanation) are encountered in most cultures, they are connected in highly complex ways and that deviations occur as a result of the specific characteristics (e.g. literacy) of a particular culture.

The influence that culture has on intellectual abilities can also assume different forms. Apart from the direct learning situations inside and outside the school, the typical behaviour code of the community also makes itself felt in subtle ways. For example, it appears that a convergent style of thinking occurs more frequently among young people who grow up in authoritarian, traditional communities than

is the case in freer communities (Guthrie 1963; Ghuman 1980).

From the above it is evident that, although the influence of culture is in many respects subtle and difficult to observe directly, it can nevertheless be an important source of bias in tests, especially in plural communities with divergent cultural backgrounds.

6.2 WHAT IS CULTURE?

> Broadly speaking, *culture* is generally conceptualized as the particular traditions, values, norms, and practices of any people who share a common ancestry ... . Assessment, especially test data, gathered by school psychologists and other practitioners is - in varying degrees - culturally shaped (Valencia & Lopez 1992: 400).

The definition given by Sue and Sue (Helms 1992: 1091) is more comprehensive: they propose that culture "consists of all those things that people have learned to do, believe, value, and enjoy in their history. It is the totality of ideals, beliefs, skills, tools, customs, and institutions into which each member of a society is born."

6.3 THE INFLUENCE OF CULTURE ON TEST PERFORMANCE

According to Anastasi (1990: 355), cultural differences may operate in various ways to bring about group differences in behaviour. This author maintains that the level at which cultural influences are manifested varies along a continuum extending from superficial and temporary effects at one pole to those that are basic, permanent and far-reaching at the other. Even relatively trivial experiential differences may have the effect that some test items are worthless for individuals from certain cultures. Broadly speaking, the same cultural factors that affect test performance are also likely to have an impact on the wider behaviour domain that the test is designed to sample.

The longer a particular environmental condition has operated in the person's lifetime, the more difficult it becomes to reverse its effects. Certain conditions that are environmentally determined are not necessarily remediable. In this regard, Anastasi (1990: 356-357) gives the following example.

> In a series of studies on large samples of blacks and whites, prenatal and perinatal disorders were found to be significantly related to mental retardation and behavior disorders in the offspring. An important source of such irregularities in the process of childbearing and birth is to be found in deficiencies of maternal nutrition and other conditions associated with low socioeconomic status. Analysis of the data revealed a much higher frequency of all such medical complications in lower than in higher socioeconomic levels, and a higher frequency among blacks than among whites. Here then is an example of cultural differentials producing organic disorders that in turn may lead to behavioral deficiencies.

Since all behaviour is affected by the cultural milieu in which the person is reared and since psychological tests are but samples of behaviour, it follows that membership of a particular cultural or ethnic group can be expected to have some effect on test scores. Jensen (1980: 127) argues that "in an intelligence test the specific content of the items is unessential, so long as it is apprehended or perceived in the same way by all persons taking the test... The content of the items is a mere vehicle for the essential elements of intelligence test items." But according to Miller-Jones (1989), it is precisely the issue of determining uniform item apprehension that is at the centre of the concern for cultural influences on testing. This author is of the opinion that it may be impossible to achieve task or context equivalence between highly divergent cultures. (Do blacks and whites in the United States, or in South Africa, constitute "highly divergent" cultures?) Thorndike *et al.* (1991) add another element to the discussion. According to them, the critical issue is the degree to which cultural factors affect the criterion behaviour. If cultural background affects test scores but not criterion scores or behaviour, then the test is undoubtedly *unfair*. On the other hand, a test is

considered more fair if cultural background affects both the test score and the criterion. These authors further point out that, even if both the test score and the criterion are affected, the question remains whether the particular test should be used at all. Answers to this question would depend on the purposes for which the test is used. "Ethical problems concerning the assessment of minorities do not stem so much from the tests themselves as from ways the tests are used and in particular the inferences that are drawn from the test scores" (Thorndike *et al.* 1991: 444).

When psychologists began to develop measuring instruments for cross-cultural testing in the first part of this century, they hoped it would be possible to measure *hereditary intellectual potential* (Anastasi 1990: 357) independently of the influence of cultural background. The instruments they produced in this regard were then called *culture-free* tests. Subsequent developments in genetics and psychology, however, have demonstrated the fallacy of this concept. "We now recognize that hereditary and environmental factors operate jointly at all stages in the organism's development and their effects are inextricably intertwined in the resulting behavior" (Anastasi 1990: 357). To try to develop an instrument that is totally free from cultural influences is therefore futile. Because the tests do not access *underlying ability*, there is no value in trying to invent a test that is universally applicable or one that is culture-free (Olson 1986). Moreover, it is also unlikely that any test can be equally "fair" to more than one cultural group, especially if the cultures are highly dissimilar. Although it is possible to reduce cultural differentials in test performance, cross-cultural tests cannot completely eliminate such differentials because every test tends to favour individuals from the culture in which it was developed.

The present objective in cross-cultural testing is rather to develop tests that presuppose only experiences that are common to the different cultures concerned. In this process, such terms as "culture-common", "culture fair" and "cross-cultural" have replaced the earlier "culture-free" (Anastasi 1990). Most traditional cross-cultural tests make use of nonverbal content in order to obtain a more culture-fair measure of intellectual abilities; the assumption is that nonverbal content

measures the same intellectual functions as the verbal tests. This assumption (that the two kinds of tests measure the same functions) is questionable, according to Anastasi (1990). An opposite view is taken by Olson (1986), but the conclusion regarding the "culture-fairness" of nonverbal content is the same: because the two kinds of contents measure the same functions, the nonverbal content is equally culturally biased! Olson illustrates his viewpoint by referring to Raven's Progressive Matrices. According to him, Raven's Matrices are highly related to literacy. Why? - because the Raven's requires the same analytical rules, rules for analysis, coding, and transforming relationships required by the analysis of verbal content. Olson is furthermore of the opinion that cultural inventions are inventions that must respect the cognitive structures of their users, otherwise they cannot be learned and used. From this perspective, Olson views the structures of cultural artifacts as explications of the mind. From the fact that "our culture and technology permit us to put a man on the moon" Olson (1986: 356), comes to the conclusion that "to attempt to characterize intelligence independently of those technologies seems to be a fundamental error" (1986: 356). Intelligence tests, however, do reflect culture specifics, i.e. they do not apply across cultural groups, and hence they correlate highly with social class and performance is dramatically affected by schooling (Olson 1986: 358).

There seems to be a growing body of evidence suggesting that nonlanguage tests may in fact be more culturally loaded than language tests (Anastasi 1990: 359). Culture loaded items are items involving pictures of cultural artifacts such as vehicles, furniture, musical instruments or household appliances, while culture-reduced items involve lines, circles, triangles and rectangles (Jensen 1980: 133). An interesting finding mentioned by Jensen (Dyck 1996: 68) is that "the average white-black difference [in performance] is *greater* on the items judged as 'least cultural' than on items judged as 'most cultural,' and this remains true when the 'most' and 'least' cultural items are equated for difficulty (percentage passing) in the white population." For this reason nonverbal tests have fared no better than verbal tests in the testing of minority groups within the United States. From a somewhat different angle, culturally reduced tests display certain limitations. In this regard it has been noted by Vernon (Wood 1986: 30) that "the further one

tries to get away from tests that are culturally conditioned, the less accurate they become as predictors of future educability." It can hardly be otherwise, because intellectual abilities are always an interaction between biological tendencies and opportunities for learning in a particular cultural context. Abilities cannot be conceptualized or measured with accuracy independent of the particular context in which the person happens to live (Gardner & Hatch 1989). Indeed, as White (1988) has pointed out, culture distributes the opportunities to exemplify intelligence unevenly. In a world without ballet there would be no Baryshnikov, or without a well-developed physics, no Einstein; someone like Bobby Fischer might have had the potential to be a great chess player, but if he had lived in a culture without chess, that potential would certainly never have been realized.

6.4 POSSIBLE SOLUTIONS TO THE PROBLEM OF CULTURAL INFLUENCE ON TESTING

As has been evident from the previous section, it is futile to try to remove cultural influence on test performance or in the development of measuring instruments. What, then, is the solution? In considering possible solutions, we are thinking in the first place of different cultural groups sharing the same territory and the same government as is the case in the United States and in South Africa.

Banning all assessment and measuring devices, as many, especially minority groups, wanted in the 1970s in the United States, is certainly one possibility. As this amounts to breaking the thermometer just because it does not register a fever with hundred per cent accuracy in certain groups, it was not a viable option in the United States nor will it be in South Africa. This avenue will therefore not be explored further.

Alteration of norms and tasks is one way of modifying psychometric practices for persons with racial/ethnic[2], cultural or language differences. The issue of subgroup norming, i.e. basing normative reference data on subgroups of a

---

[2] In accordance with, among others, Singham (1995), Dyck (1996) and Moore (1987), preference is given to the term "ethnicity", which is a much more meaningful concept for understanding black-white differences than "race", which has biological overtones.

population rather than on the total group, has been hotly debated in employment testing in the United States for many years. The controversy over subgroup norming reached a new peak with the passage of the Civil Rights Act of 1991, which banned any form of "score adjustment on the basis of race, colour, religion, sex or national origin" (Brown 1994: 927).

Score adjustment takes a number of forms, including correction for imperfect prediction, adding a fixed number of points to the scores of particular groups (bonus points), within- group norming (separate norms), top-down selection from separate lists, sliding bands and minority preference.

6.4.1 Correction for imperfect prediction

This correction adds points to minority test scores so that minority applicants and majority applicants who would perform equally well if selected would have the same adjusted test scores. The adjustment is achieved by adding $(1-r^2)m$ points to each minority test score and then selecting in order of the adjusted scores; r is the correlation between test scores and job performance and m is the difference between majority and minority test means (Kehoe & Tenopyr 1994: 297).

6.4.2 Bonus points

Certainly the most direct form of score adjustment involves adding a fixed number of points to the scores of all individuals who are members of a particular group. The goal is the reduction or elimination of differences between certain groups. If, for example, there is a 15-point mean difference between blacks and whites on a particular test, the mean difference can be eliminated by adding 15 points to the scores of all members of the lower scoring group. The effectiveness of this procedure in eliminating *adverse impact* is, however, dependent on the comparability of test standard deviations across groups. If the lower scoring group has a smaller standard deviation than does the higher scoring group, adverse impact is still possible even after score adjustment to eliminate mean differences (Sackett & Wilk 1994: 936).

6.4.3 Within-group norming

Within-group norming or separate norms involves converting individual scores to either standard scores or percentile scores within one's group (Sackett & Wilk 1994: 937). This approach is sensitive to differences in both means and standard deviations across groups, and is therefore more successful in reducing adverse impact than is the bonus point approach. When a given score has a markedly different psychological meaning in one group than in another, e.g. male and female who respond differently on an interest inventory, the appropriate score for meaningful interpretation is one's standing within the group. To make provision for certain gender differences by means of separate norms is generally acceptable, but the idea of "race norming" for different ethnic groups seems to be "mired in controversy" (Helms 1992: 1083). An example of the questions that are being asked in this regard is given by Brown (1994: 928). Two individuals, one white and one black, obtained the same adjusted percentile score on the General Aptitude Test Battery (GATB). "For the adjusted 70th percentile, a White individual scored 327 and an African American individual scored 283. How do we answer the question, Why did these two individuals earn the same percentile score? Is subgroup norming a legitimate, empirically supported method of reducing the adverse impact on protected groups of the use of selection tests? Or is it a social agenda, paramount to preferential treatment, that found a rationale for itself within statistics?"

6.4.4 Top-down selection from separate lists

This approach involves ranking individuals separately within groups and then selecting top-down from within each group in accordance with some preset rule as to the number of vacancies that will be allotted to each group (Sackett & Wilk 1994: 937). When the allocation rule is determined by group representation in the applicant pool, this approach is the same as within-group norming. If there are 10 vacancies, the top five whites and the top five blacks would be selected.

6.4.5 Separate cutoffs

The use of separate cutoffs for different groups is in practice the same as the

bonus point approach: setting a cutoff 10 points lower for one group produces the same outcome as adding 10 points to the scores of members of that group. Separate cutoffs makes it very clear that a lower standard is being used for one group than for another (Sackett & Wilk 1994: 937).

6.4.6 Sliding bands

One motivating factor for the sliding band approach is increased minority selection. The following example explains how the sliding bands approach functions in practice (Sackett & Wilk 1994: 938; Kehoe & Tenopyr 1994: 297). Assume that the top score in a particular test is 100 and the first band includes scores from 91 to 100:

(i) select minority group members top-down within the first band; select majority group members scoring 100. Once all individuals scoring 100 have been selected, the highest raw score remaining is 99,

(ii) slide the band from 91-100 to 90-99 and select minority group members scoring 90; select majority group members scoring 99,

(iii) slide the band from 90-99 to 89-98 and select minority group members scoring 89; select majority group members scoring 98,

(iv) continue in this fashion until vacancies are filled.

From the above it is evident that this approach is exactly the same as a bonus point approach, with band width as the number of points to be added to the scores of minority group members. This approach can also be seen as producing different cutoffs. Without a minority preference component, the sliding bands approach generally has little impact on the rate of minority group selection. Viewed from one angle, score adjustment is an attempt to reduce cultural and other influences on testing for personnel selection. From another angle, score adjustment is a practical approach to accommodate the (mistaken) belief that group differences *per se* indicate a flaw in the measuring device. Be that as it may, the function of score

adjustment is to introduce bias against the higher scoring group (usually whites in the case of the United States) in measuring job-related skills and abilities. For example, a white individual would have to score around the 84th percentile on the GATB to have the same chance of being accepted as a black individual scoring at only the 50th percentile for whites. Because of this, Gottfredson (1994: 957) argues that "race-norming is destructive social policy because, among other side effects, *it would make permanent the very social inequalities it is supposedly intended to eliminate*" (emphasis added). She further argues that (1994: 963)

> personnel-selection psychology can also perform an important service by analyzing the full panoply of costs and benefits of different strategies for reducing disparate impact. But the biggest contribution personnel psychology can make in the long run may be to insist collectively and candidly that *their measurement tools are neither the cause of nor the cure for racial differences in job skills* and consequently inequalities in employment (emphasis added).

6.4.7 A Eurocentric versus an Afrocentric approach to testing

Helms (1992) argues that cognitive ability tests have been constructed on the basis of Eurocentric values, which are different from Afrocentric values. She defines Eurocentricism as "a perceptual set in which European and/or European American values, customs, traditions and characteristics are used as exclusive standards against which people and events in the world are evaluated and perceived" (1992: 1093). According to Helms, there are a number of values and beliefs of the Eurocentric worldview that may have particular relevance to the area of test construction and validation. Of these, the following may be the most harmful to other cultural groups: (i) dualistic linear or rational thinking, (ii) the White superiority assumption and (iii) the emphasis on the scientific method for discovering intellectual ability. Each of these values may influence test construction, testing procedures and test interpretation.

With regard to Afrocentricism, Boykin and Toms (Helms 1992: 1096) have

proposed various dimensions of African culture that might be applicable to the testing process. Helms summarizes eight of these dimensions:

(i) Spirituality - greater validity of the power of immaterial forces in everyday life over factual thinking.

(ii) Harmony - the self and one's surroundings are interconnected.

(iii) Movement - personal conduct is organized through movement.

(iv) Affect - integration of feelings with thoughts and actions.

(v) Communalism - valuing of one's group more than individuals.

(vi) Expressive - unique personality is expressed through one's behavioural style.

(vii) Orality - knowledge may be gained and transmitted orally.

(viii) Social time - time is measured by socially meaningful events and customs.

It is important to note that these authors are not proposing a nonintellectual form of intelligence; rather they are asserting that efficient use of African-centred cognitive abilities requires awareness and integration of social contextual factors into one's thinking process. This means, according to Helms (1992: 1096), that African-centred information-processing strategies might be implicit unmeasured aspects of cognitive ability tests as well as the criteria these tests are used to predict.

Heath (Helms 1992: 1097) argues that

> from childhood, Black Americans are socialized in Black communities to develop spontaneous, creative, interactive, and expansive thinking skills. Consequently, upon reaching testable age, it is difficult for

them to reconcile the contrasting socially oriented worldviews of their communities with the ascetic Eurocentric view that presumably underlies test construction, particularly when they are bombarded with information to the effect that test scores and intelligence are synonymous.

Helms (1992) suggests that existing tests be modified to include greater cultural variety and that new types of cognitive assessment be developed and standardized. The inclusion of Black African-American culture in cognitive ability assessment procedures should result in fairer assessment of Black Americans' general cognitive ability levels. These ideas, together with those of Geisinger (1994) concerning the adaptation of a measuring instrument from an original culture to a new one, may prove useful for psychometricians working in multicultural societies.

In order to address the need for tests that are reliable indicators of performance of persons from predominantly "non-mainstream-western", "non-middle-class" backgrounds, Davidson (1995) proposed a multiaxial model of cognitive assessment that could be used to assess indigenous Australians. This model is based on indigenous and everyday judgments about cognitive performance which can replace psychometric testing of indigenous Australians.

In the development of cognitive behavioural scales and cognitive assessments there is considerable potential for, and value in, constructing a cognitive demand axis whereby performance is assessed in relation to the complexity of cognitive functions involved, and the familiarity and perceived degree of difficulty of the task. Such an axis might also include an acculturation quotient, as Helms (1992) has suggested. In addition, a Global Assessment of Cognitive Functioning scale measuring everyday-life behaviour from cognitive competence and social acceptability in a wide range of everyday-life contexts to inability to function in most basic everyday-life contexts would provide additional valuable information about cognitive

performance on criterial tasks or in criterial contexts (Davidson 1995: 33).

Although most cross-cultural psychologists would probably not question whether Davidson's model or Helm's suggestions might be an effective way of gaining important information about an individual's cognitive function, many of them would certainly question the necessity of developing an assessment model specifically for one cultural group as opposed to others sharing the same country. In this regard, Dyck's (1996: 66) reaction to Davidson's model is significant: "I argue that such a racially specific approach to assessment is based on inappropriate racial stereotyping, a confounding of cultural (categorical) variables with individual differences (continuous) variables, and a misrepresentation of evidence on cultural bias in cognitive abilities tests". And further: "What is controversial is the idea that a special or unique model of assessment must be created for assessing the cognitive functioning of indigenous Australians as opposed to all other Australians". Dyck comes to the conclusion that "To argue that indigenous Australians are so different from all other Australians that their cognitive functions must be assessed in a unique way could be taken as racist". Dyck's answer to the question of whether standard ability tests are culturally biased is in accordance with that of many other authors (e.g. Jensen 1980; Brown 1994; Gottfredson 1994; Sackett & Wilk 1994): these tests are not biased. Dyck (1996: 68) suggests that "it is time to stop blaming 'test bias' for the lower average performance of indigenous Australians on many cognitive ability indicators and ask what conditions are responsible for the lower average performance".

In Davidson's (1996: 71) reply to Dyck's (1996) criticism of Davidson's (1995) proposal for a multiaxial model of indigenous cognitive assessment, he points out that "tests can be unbiased in a statistical sense, but unfair in a cultural sense, in that disproportionate numbers of minority and majority culture members are selected for a particular purpose". This illustrates the whole dilemma surrounding the question whether standardized tests are biased or not. In the strict sense of the term *bias*, Jensen, Dyck and others are correct (tests are by and large not biased), but those such as Helms and Davidson who maintain that tests can be

statistically *un*biased and yet at the same time be *culturally biased* in the sense that the tests are culturally unfair have an equally valid argument. Thus, on the one hand there are those who follow a *Eurocentric* approach to test development and believe that if there is any bias in a test, it can be detected by statistical and other techniques; on the other hand, there are those who believe in the same statistical techniques but at the same time maintain that these techniques have failed to detect *Eurocentric bias* and consequently suggest an alternative, culturocentric approach to the development of cognitive tests.

This dilemma would appear not easily soluble, neither in the United States nor in South Africa, where the supposedly Eurocentric approach to testing is increasingly being questioned. Why? Because there is a crucial element lacking in the debate surrounding a choice between a Eurocentric and a culturocentric/Afrocentric approach to test development. That element is culture itself. In spite of all that has been said about the influence of culture on test performance, very little empirical evidence is available on the effects of specific cultural practices. The problem in this regard is that specific cultural practices are seldom incorporated as dependent variables in experiments. The Laboratory of Comparative Human Cognition (1979: 168-169) has the following to say in this connection:

> Culture is still distressingly absent on the dependent variable side of a great deal of cross-cultural work .... The absence of well-defined theories of the task-specific activities which give rise to the dependent variables is a central source of the ambiguity in almost all this work. ... . Cases in which there is a strong theory of the task and its relation to cultural practices point the way to incorporating culture into our dependent variables. As cultural practices become the focus of more and more cross-cultural cognitive work, greater emphasis will have to be put on developing cognitive ethnographies which go beyond cognitive anthropology's current products. A new concern for specifying culturally organized activities on a level which the psychologist can use is one of the major tasks confronting the study of culture and cognition in the coming decade.

The suggestions of Helms, Davidson and others regarding a culturocentric/Afrocentric approach to testing are interesting, but the necessary theory on which test development can be based has not yet been developed. "The major methodological lesson", according to the Laboratory of Comparative Human Cognition (1979: 164), "is that ethnographic analysis of cultural activities that require and promote particular cognitive skills must be carried out in close proximity with (and preferably prior to) experimental analysis of the skills in test-like situations. Otherwise, we remain critically ignorant of how behaviors sampled in the test relate to those routinely demanded by the culture."

A way out of the impasse reached between a Eurocentric and an Afrocentric approach to testing and assessment (at least on an educational level) - one that also holds promise for testing in South Africa - has been suggested by Valencia and Lopez (1992). These authors maintain that the focus of assessment in schools should be the *school culture*. If one does not choose an answer to the question, "In reference to what culture am I assessing a student's degree of adequate or inadequate functioning?" (1992: 416-417), factors important to deciding such issues as whether a student's low achievement is the result of environmental, cultural or economic disadvantage will probably be disregarded.

The culture to which psychoeducational assessment refers must, in the opinion of Valencia and Lopez (1992: 417), primarily relate to the school culture. The authors support their viewpoint as follows:

> In this respect, although it is essential to consider a student's home culture to determine the effects on school functioning, adaptation to the school culture is the primary issue of eligibility for special education. It is true that special education conditions, such as mental retardation and serious emotional disturbance, must, according to their definitions, be manifest in the home setting as well as at school, *yet they must be clearly evident at school to be relevant to special education* (emphasis added).

According to Valencia and Lopez (1992: 417), identifying the referent culture of psychoeducational assessment as the school culture has the advantages that it

- puts assessment into a realistic and more manageable context
- restricts such controversial activities as labelling to the school setting (these labels are relevant to the school setting and should preferably not be used outside of that setting)
- does not minimize the significance of cultural differences

Looking at the school setting as the referent permits every aspect of the assessment process to be evaluated by the question, "What relevance does this activity have to the student's adaptation to the school culture?" (1992: 417). This question gives psychoeducational assessment a cross-cultural orientation.

> Assessment can be cross-cultural because of the common core of educational objectives held for all students. Otherwise ... the only way to assess the burgeoning racial and ethnic minority school population in the United States is to have entirely different assessments for each group - an impossible task. This does not imply, however, that schools should neglect the unique instructional needs of culturally and linguistically diverse minority children (1992: 418).

In conclusion, we in South Africa, with our different cultural and ethnic groups, can, for obvious reasons, benefit immensely by the viewpoints put forward by Valencia and Lopez. Whatever the attitudes and views of the education authorities may be, if South Africa wants to continue in its role as a significant player on international markets, the educational objectives set for the country cannot be vastly different from those of our trading partners. By emphasizing and promoting a common core of educational objectives and identifying the referent culture as the *school culture*, psychoeducational testing and assessment in South Africa can be cross-cultural. Consequently, common measuring instruments, based primarily on a Eurocentric approach, can be used with the various groups. Otherwise, different

measuring instruments and assessment techniques must be developed for each cultural group - which is not only impractical but also economically unaffordable.

7. THE ROLE OF PSYCHOLOGICAL TESTS IN SOUTH AFRICAN SCHOOLS

7.1 CLASSIFICATION OF PSYCHOLOGICAL TESTS

Psychological tests may be classified or characterized in many different ways, for example by their content area, intended uses, method of test administration, strategy followed in item construction, type of stimuli or responses, test interpretation, standardization, criteria for scoring and so on. The most common method, however, is to classify tests by their content or by the attributes they measure, for example, musical ability, mechanical aptitude, spatial ability, scholastic aptitude, school readiness or personality traits.

Since psychology is mainly concerned with two broad categories of human behaviour, *cognitive* and *affective*, tests are usually classified accordingly.

Cognitive measures

- Individual Intelligence Scales - mainly for clinical purposes
- Group Intelligence Tests - preliminary screening instruments, to be followed by tests of special abilities or aptitudes
- Multiple Aptitude Tests - focus on potential or future behaviour
- Achievement Tests - measure development and learning to date

Affective measures

- Personality Tests - designed to measure relatively stable traits
- Interest Inventories - measure a person's preferences and aversions
- Attitude Scales - assess the individual's predisposition to think, feel and behave toward a particular social object
- Adjustment Scales - measure behaviour patterns concerning one's adjustment to the immediate environment

7.2 TEST DEVELOPMENT IN SOUTH AFRICA

Most countries in the world have some or other organization responsible for the development of psychometric instruments which can be used to advance the economic, social and educational welfare of their people.

The Psychological Test Divisions of the HSRC were brought into being with the express purpose of developing tests in a South African context. Psychological measurement is founded on the well-established concept that there are universal human characteristics and abilities. Therefore the approach to the measurement of characteristics/capabilities should not radically depart from approaches used throughout the world.

The HSRC has developed a wide range of products which may be used for a variety of purposes in schools, tertiary institutions and the private sector. At present (1997) the HSRC supports about 60 test batteries, individual intelligence scales, personality questionnaires and interest questionnaires. For example the (see HSRC Test Catalogue):

Academic Aptitude Test (AAT)
Aptitude Tests for School Beginners (ASB)
Scholastic Aptitude Test Battery (SATB)
Technical Aptitude Test Battery for Low Literates (TAB)
Trade Aptitude Test Battery (TRAT)
Individual Scale for General Scholastic Aptitude (IGSA)
South African Individual Scales (SSAIS-R)
Individual Scale for Xhosa-speaking Pupils
Jung Personality Questionnaire (JPQ)
High School Personality Questionnaire (HSPQ)
19 Field Interest Inventory (19 FII)
Survey of Study Habits and Attitudes

That South Africa is not way off the mark as far as test development and usage

are concerned is evident from the information supplied by Thomas Oakland (1995), Professor of Educational Psychology at the University of Texas, who conducted a 44-country survey of the usage of both domestically developed and imported tests for children and young people. He found that testing is most commonly used for diagnostic purposes, as well as for guidance, admissions and placement purposes. Intelligence and personality tests predominate. Schools and clinics are the most common testing sites. A very important finding is that the test use patterns are remarkably similar for highly industrialized, less industrialized, developing third world and Middle East countries - the mean number of tests used in these four groupings ranges from 16 to 19. On the other hand, only about seven tests are used in the least developed bloc of countries (where very few resources for developing tests exist). Test development and usage in South Africa clearly fits the international pattern set by the four groupings mentioned above.

There are approximately 11 million pupils at school who will eventually need, for example, guidance regarding subject and career choices; some of the pupils (about 10%) will need assistance regarding learning problems. In all these instances measuring instruments are needed to provide the information for decision-making. Good quality information is essential for making informed and responsible decisions. These decisions may have to be made by the individual student, the parent, the counsellor, the principal and the education department. Instruments that are needed for decision-making include aptitude tests (various cognitive abilities), group intelligence tests, individual intelligence scales, personality tests and questionnaires and interest questionnaires.

The measuring instruments listed in the HSRC Test Catalogue can contribute to efforts directed at meeting the challenges facing the "new" South Africa in many ways. These instruments are especially suited for subject and career guidance at school, for the diagnosis of learning problems and for the selection and placement of persons in appropriate jobs.

Some of the instruments that are extensively used in schools date back to the 1970s. A need has therefore been identified for revamping these instruments. A

further shortcoming is that the norms of many of the existing instruments are not applicable to the total South African population and need to be revised.

7.3 COGNITIVE TESTS DEVELOPED BY THE HSRC

The cognitive tests developed by the HSRC include individual intelligence scales, group intelligence tests, and aptitude and proficiency tests.

7.3.1 Individual intelligence scales

The following scales are available:

- Individual Scale for General Scholastic Aptitude (ISGSA)
- Individual Scale for Northern Sotho-speaking pupils
- Individual Scale for Southern Sotho-speaking pupils
- Individual Scale for Tswana-speaking pupils
- Individual Scale for Xhosa-speaking pupils
- Individual Scale for Zulu-speaking pupils
- Junior South African Individual Scales (JSAIS)
- Senior South African Individual Scale - Revised (SSAIS-R)
- South African Individual Scale for the Blind (SAISB)

Uses of individual intelligence scales

Individual testing allows the tester to observe directly the behaviour of the testee in the test situation, which is usually a valuable additional source of clinical information that is not reflected in the numerical score. No reading is required on the part of the testee, so it is possible to test young children and people of limited literacy.

The most important function of individual intelligence tests is usually to measure the general intelligence factor, but with the emphasis on those facets of intelligence that are closely related to efficient functioning in the modern

technological milieu. The assumption is usually made that the total score of subtests in the intelligence scale represents an underlying general factor of intelligence (Spearman's g factor). An individual scale is used for diagnosing different levels of mental retardation with the further aim of providing different levels of special education, training and care for the persons involved.

Another important purpose of individual intelligence scales is usually to provide scores for as many as possible of the mental abilities that are related to intelligence. Intelligence must always be seen in relation to the life phase in which evaluated individuals find themselves. At school it would therefore involve those mental abilities that are important for scholastic achievement. The purpose of individual intelligence testing is therefore to obtain a profile of the strong and weak points of a testee's intellectual functioning. For instance, information may be obtained on a testee's ability to handle words and symbols and the ability to manipulate objects or to observe visual patterns. Differential achievement in verbal and performance scales often throws light on the nature of specific learning problems and may sometimes even indicate the existence of pathological conditions (cf. Hay & Pieters 1994). The study of a profile of test scores provides useful information on how the test results for individual subtests can be interpreted. Additional information is, however, necessary to support or reject the original diagnosis.

Individual intelligence tests were developed before group tests and are almost always included in any comprehensive psychological assessment that involves testing. Individual tests are essential for identifying people who need remedial intervention (Seligman 1994). Individual tests form the backbone of school and clinical psychology practice, student counselling, private practice and mental health institutions.

An advantage of individual intelligence tests over group tests is that the psychologist can observe whether testees are really trying their best or, if that is not possible, at least know that the testees were not fully engaged in trying to answer the test questions. Thus poor scores on group tests may sometimes be

due to the fact that testees were not motivated, or were actively trying to do badly. "For all these reasons *if the most accurate assessment of intelligence is required an individual intelligence test should be given*" (Kline 1993: 186) (emphasis added).

An individual intelligence test may occasionally be used productively outside the target group for which it was originally intended. Consider a case in which a mentally retarded 19-year-old individual is tested with the JSAIS and is found to have a test age of five years. This information may be most useful to the psychologist in planning further training and care needs for the person. In this example the JSAIS should not be seen as an IQ test but rather as an achievement test for certain cognitive tasks. This example also illustrates that, if the most accurate assessment of intelligence is required, an individual intelligence test should be included in the assessment process.

Another example where a test may be considered for use outside the target group is the SSAIS-R for individuals who understand Afrikaans or English reasonably well, but do not have one of these languages as a home language. Van Eeden (1993) investigated the applicability of the SSAIS-R for children who have an African language as mother tongue and who are in private English-medium schools. The study indicated that, although the factorial structure differed somewhat from the norm group, the SSAIS-R total score and subtest scores predicted scholastic achievement equally well for English and non-English speakers.

To conclude, an individual intelligence test is indispensable for an accurate assessment of intelligence. Within the context of the school culture, an instrument of this type can provide valid results for various cultural groups.

7.3.2 Group intelligence tests

The following tests are available:

- General Scholastic Aptitude Test (Junior)

- General Scholastic Aptitude Test (Intermediate)
- General Scholastic Aptitude Test (Senior)
- SA Group Test for Partially Sighted Pupils
- Group Tests for 5/6 and 7/8 year-olds
- Mental Alertness (Intermediate)
- Mental Alertness (Advanced)
- High Level Figures Classification
- Figure Classification Test
- Conceptual Reasoning Test
- Deductive Reasoning Test

Uses of group intelligence tests

In group intelligence testing it is generally assumed that the ability to solve problems with regard to figures, verbal material (words and sentences) and numbers is an important predictor of those facets of intelligence that are important in normal mental functioning in a technological milieu. It is further assumed that the subtests (and items) provide a measure of Spearman's g factor of intelligence. Because some people may show marked differences with regard to their ability to solve problems with verbal and non-verbal content, tests usually consist of 50 per cent verbal and 50 per cent non-verbal items.

The most obvious advantage of group intelligence tests is that it is possible to test a large number of persons at once, something which is essential for any large-scale testing programme. Another advantage is that the person administering the test does not require the same level of skill and training as does the psychologist administering an individual intelligence test.

Group intelligence tests for use at school are designed to measure academic intelligence, in other words, scholastic ability. The tests can be used as objective aids to determine pupils' reasoning ability or problem-solving ability. This information, together with biographical and other data, can be used to design optimal teaching strategies for pupils. The tests can also be used to identify those

pupils who will need more time-consuming individual testing and remediation. A group test score provides a measure of what the person can do at the time of testing.

Discrepancies between pupils' achievement at school and their general intellectual ability, as indicated by a group intelligence test, seem fairly general. Although there may be various reasons for this, the important point is that at least some of these problems can be eliminated by remedial actions. Without a group intelligence test it would be very difficult to determine whether poor school achievement can be ascribed to low ability or to under-achievement. In other words, with the aid of a group intelligence test a distinction can be made between poor school performance due to lack of ability and poor performance due to reasons that have very little to do with intellectual ability.

According to Seligman (1994: 132-133), intelligence tests are especially useful in the following instances:

1) They can facilitate selection of students for gifted and talented programs or other programs offering the opportunity for advanced or accelerated course work.

2) They can provide a measure of functioning that is less linked to educational experiences than most achievement tests and many aptitude tests and, therefore, provide a different source of information on abilities.

3) A disparity between school performance and inventoried intelligence can be helpful in identifying children who are performing below capacity as well as those who are stretch-ing their abilities and may feel great pressure to achieve academic success.

4) Intelligence test scores correlate significantly and positively with many variables relevant to career development such as occupational success, career maturity, levels of occupational aspiration, academic performance, and likelihood of attending and graduating from college. Most people are aware of their intellectual abilities and gravitate toward career paths that are consistent with those abilities.

Seligman (1994: 133) also points out that the patterns of use of intelligence tests have changed greatly since the 1970s: "Schoolwide testing of intelligence now has all but disappeared, and most testing is done with individuals or small groups, with the specific purpose of the testing predetermined." Although counsellors make less use of intelligence tests than they do of other types of tests, Seligman recommends that they should nevertheless be comfortable with intelligence tests, not only for their own use, but for understanding psychological reports and especially for knowing when a referral for intelligence testing is warranted.

In view of the fact that group intelligence tests are closely related to both aptitude and achievement tests, and individual's scores on all three types of instruments tend to be highly correlated, counsellors should be sure that they are not administering an intelligence test when what they actually want is an aptitude test. The latter type of test is usually more appropriate for prediction. It should also be borne in mind that both intelligence and aptitude tests correlate more highly with success in training than with success on the job, and intelligence tests, like aptitude tests, are not good indicators of overall career success or satisfaction (Seligman 1994).

In interpreting group intelligence test scores other information (e.g. level of motivation, background and academic record as well as other test data) should also be taken into account. While it is generally true that intelligence tests may be useful in estimating the chances of success of individuals in a particular educational or occupational endeavour, it is also true that "intelligence tests do not reflect

innate ability or true intellectual capacity" (Seligman 1994: 134).

To conclude, for most purposes group intelligence tests are satisfactory. Where there is some specific problem with a person, however, then an individual intelligence test is to be preferred (Kline 1993).

7.3.3 Aptitude and proficiency tests

The following *aptitude* tests or batteries are available:

- Aptitude Tests for School Beginners (ASB)
- Aptitude Test Battery for Adults (AA)
- Senior Aptitude Tests (SAT)
- Senior Aptitude Tests for Partially Sighted Persons (SAT-S)
- Junior Aptitude Tests (JAT)
- Senior Academic-Technical Aptitude Tests (SATA)
- Academic-Technical Aptitude Tests (ATA)
- High Level Battery
- Intermediate Battery
- Normal Battery
- Senior Musical Aptitude Test (MUSAT S)
- Junior Musical Aptitude Test (MUSAT J)
- Aptitude for and Sensitivity to Music - Senior Test (ASM S)
- Aptitude for and Sensitivity to Music - Junior Test (ASM J)
- Programmer Aptitude Battery (PAB)
- Trade Aptitude Test Battery (TRAT)
- Technical Aptitude Test Battery for Low Literates (TAB)
- Industrial Test Battery (ITB)

The following *proficiency* tests or batteries are available:

- Tests for Oral Language Production (TOLP)
- Scholastic Aptitude Test Battery for Pupils in Stds 2 and 3 (SATB Stds 2/3)

- Scholastic Aptitude Test Battery for Pupils in Stds 4 and 5 (SATB Stds 4/5)
- Scholastic Aptitude Test Battery for Pupils in Stds 6 and 7 (SATB Stds 6/7)
- Guidance Test Battery for Secondary Pupils (GBS)
- Academic Aptitude Test for Pupils in Std 10 (AAT Std 10)
- Academic Aptitude Test for University Students (AAT Univ.)
- High Level Estimation Test (ET-HL)
- Standard Level Arithmetic Reasoning Test (ART-SL)
- High Level Arithmetic Reasoning Test (ART-HL)

Uses of aptitude and proficiency tests

Any test, according to Bingham (1937), is a test of aptitude insofar as the score gives an indication of future potentialities. *Predictive value* is therefore the most characteristic feature of an aptitude test; without it a test is not an aptitude test. By using an aptitude test we wish to determine whether a person *now* has the ability to carry out a certain task in the *future*, if he or she receives the necessary training in the intervening period. In other words we wish to determine whether a person has the necessary *learning ability* in a specific direction to enable him or her to achieve success in that direction if appropriate stimuli are provided. According to Seligman (1994: 119),

> Aptitude tests are designed to predict a person's ability to learn or profit from an educational experience or the likelihood of a person's success in a given occupation or course of study. Although, generally, achievement is developed quite rapidly, aptitude grows slowly and results from daily living and learning.

Proficiency tests, on the other hand, measure how effectively a person has utilized aptitudes and learning opportunities to gain proficiency in a particular field of study or knowledge. Proficiency tests are usually compiled in such a manner that broad educational background is tested without limiting the test compiler to syllabus content and without avoiding it completely. Proficiency tests measure, *inter alia*, knowledge and skills which were not necessarily acquired at school. Although one

can distinguish between aptitude and proficiency tests, it is generally accepted that the two types of tests necessarily overlap to some extent. For this reason the two types of tests are taken together for the purpose of this discussion about their uses.

Multiple aptitude tests, in contrast to general aptitude tests (i.e. intelligence tests), have a differential approach to the measurement of aptitude. The term *aptitude* is used here as a synonym for specific mental ability, as opposed to general mental ability (i.e. intelligence). In the light of the results of factor analyses, the term aptitude can also be associated with the concepts *group mental factor* (Vernon's model) and *primary mental ability* (Thurstone's model). Multiple aptitude tests do not provide a single or total score such as an IQ, but rather a set of scores in respect of different aptitudes. With the help of these scores an intellectual profile - showing the individual's characteristic strong and weak points - can be drawn.

The use of aptitude tests is based on the assumption that all the testees have more or less the same experience regarding the characteristics measured. If some testees have a great deal of experience in a specific area which can influence their test scores significantly, the counsellor will have to take this into consideration in the interpretation of their scores. Under such circumstances the test scores may be a reflection of skill rather than aptitude. Only if all the testees have roughly the same experience can any meaningful conclusions be drawn about interindividual differences (i.e. differences between individuals).

Aptitude tests are most commonly used for school guidance and career counselling. In other words, they are used to help people gain a greater understanding of their potential in order to facilitate decision making regarding school and career planning. The scores obtained from aptitude tests should be regarded as useful pieces of information that can be used with other information about a person in order to make certain decisions. By "other information" is meant school examination marks, interests and attitudes, study habits, hobbies, human relations, particular likes and dislikes, and so on.

Counsellors make extensive use of aptitude tests to help people decide whether they have the potential needed for specified educational or occupational goals. It should be stressed, however, that aptitude tests are not the "decision maker" but that they provide important information on the basis of which the pupil or student - in consultation with parents, teacher and counsellor - can reach realistic and judicious decisions on, for example, subject or occupational choices.

Aptitude tests do not indicate a specific occupation or provide specific answers to specific questions such as "Should this testee become an engineer?" However, aptitude tests, together with other information, can help find answers to such general questions as, "Should I go into an accountancy rather than a science direction at school? Can I consider dentistry as an occupation? Which is a more realistic choice for me: an occupation where I work with my hands or an occupation where I do a lot of thinking?" Information on a person's aptitudes is therefore essential to help him or her make realistic and considered decisions on the future.

Although aptitude test scores can be excellent predictors of school grades, correlations between aptitude tests and career success and satisfaction have not been very high (Seligman 1994). One of the reasons for this is that preparation for entry into an occupation and subsequent performance in that occupation often require somewhat different abilities. In the field of medicine, for example, success in preparation depends largely on mastery of academic courses, whereas success in performance depends also on interpersonal skills and business know-how.

To a counsellor, people's interests are just as important as their aptitudes and the two aspects cannot be viewed in isolation when considering an appropriate career choice. Generally, interests will have the greatest effect on career choice while aptitudes will be the major determinant of success in that career (Seligman 1994).

In conclusion, it must be emphasized that aptitude and proficiency tests are particularly useful in preventing wastage of talent among young people in that persons with exceptional abilities are identified at an early stage. Certainly, no

education department can do without aptitude and proficiency tests if it wants to develop the potential of young people.

7.4 AFFECTIVE MEASURES DEVELOPED/ADAPTED BY THE HSRC

The affective measures developed or adapted by the HSRC include personality tests and questionnaires and interest questionnaires.

7.4.1 Personality tests and questionnaires

The following measures are available:

- 16 Personality Factor Questionnaire (16PF)
- Children's Personality Questionnaire (CPQ)
- Clinical Analysis Questionnaire (CAQ)
- High School Personality Questionnaire (HSPQ)
- Interpersonal Relations Questionnaire (IRQ)
- Intra and Interpersonal Relations Scale (IIRS)
- IPAT Anxiety Scale
- Jung Personality Questionnaire (JPQ)
- Personal, Home, Social and Formal Relations Questionnaire (PHSF)
- Picture Motivation Tests (PMT)
- Sexual Adaptation and Functioning Test (SAFT)
- Structured-Objective Rorschach Test (SORT)
- Survey of Study Habits and Attitudes (SSHA)
- Thematic Apperception Test (TAT Z)

Uses of personality tests and questionnaires

To many psychologists, personality is just as important as interests and abilities for success in learning activities and career development. A shy and withdrawn person will probably derive as little satisfaction from a job as public relations officer in a large business as will a creative and outward-going person from a routine and

monotonous clerical job. With the aid of personality measurements, mistakes regarding career choices can in many cases be avoided. Although it is generally accepted that personality is part of career counselling, Seligman (1994: 151) points out that "the research on the relationship between personality and career development gives little clear direction as to how to explore personality and its impact on career development".

Depending on the purpose for which the test was developed, a personality test measures certain constructs that have usually been identified on a theoretical basis. Through the specific formulation of questions, constructs such as introversion-extraversion and dominance-subjection can be incorporated into a personality questionnaire.

In the same way projection tests can measure different constructs through the specific design of the stimulus material. For example, a picture of a man and a woman will elicit responses describing a man-woman relationship from most respondents.

The cards or pictures of a projection test developed for clinical purposes will include constructs such as attitude towards authority, recognition and channelling of aggression, sense of responsibility and leadership. Projection tests developed for use on children often use animal figures such as bears, rabbits and cats to measure personality traits such as parent dependence, fear of or liking for school and sociability.

Generally speaking, it can be said that structuring is the key concept in personality measurement. Practical situations continually require the evaluation of personality characteristics or traits or the prediction of behaviour arising from personality traits. Although the demonstration of the validity of any personality test is difficult because of the nature of the variables involved (Kline 1993: 217), reasonably reliable conclusions can nevertheless be drawn with regard to human functioning. In projection tests the emphasis is on measuring the more unconscious and dynamic aspects of personality. These tests provide information - which can be

used in therapy - on the interaction between forces (feelings, attitudes, etc.) within the individual that lead to specific behaviour.

Thanks to standardized and scientifically developed personality tests, employers, clinical experts, counsellors, teachers and others can, in the relatively chaotic pool of behavioural expressions, find the structuring that enables them to categorize people and predict their future behaviour. The objectives of such evaluations may include screening, classification, promotion, placement or aid with regard to adjustment problems.

To conclude, these instruments can be of considerable value for counselling, clinical and research purposes, provided they are administered in a proficient manner and the interpretation is done with the necessary care.

7.4.2 Interest questionnaires

The following questionnaires are available:

- 19 Field Interest Inventory (19FII)
- Career Development Questionnaire (CDQ)
- High School Interest Questionnaire (HSIQ)
- Life Role Inventory (LRI)
- Picture Vocational Interest Questionnaire for Adults (PVI)
- Self-Directed Search Questionnaire (SDS)
- South African Vocational Interest Inventory (SAVII)
- Values Scale (VS)
- Vocational Interest Questionnaire for Pupils in Stds 6 to 10 (VIQ)

Uses of interest questionnaires

Psychological tests are used fairly generally to achieve effective vocational and study counselling. Because of the complex composition of the human personality, it is not possible to use a single test for this purpose, so a variety of tests,

including aptitude tests, personality tests and interest questionnaires, are usually applied in attempts to obtain as comprehensive a picture as possible of a testee's cognitive and non-cognitive behaviour. Although each of these tests makes an important contribution to the success of a vocational and study counselling programme, test users are inclined to interpret the results of certain tests too simplistically. This happens with interest questionnaires, for example, on account of their relative simplicity - in spite of the fact that *interest* is no simple concept. The effectiveness of a vocational and/or study counselling programme suffering from this deficiency is questionable.

Despite the fact that *interest* is a generally known concept, there is as yet no real agreement on its psychological meaning. The many efforts made by researchers to link interest to *inter alia* attention, motivation, attitude and effect are clear evidence of this. Divergent definitions can and should therefore be expected. However, for the purposes of this discussion, the following definition, which can be found in HSRC test manuals, will suffice: "Interest is an aspect of personality and can be defined as a spontaneous attraction to, or preference for, certain activities, as well as a spontaneous aversion to other activities."

Interests involve likes and dislikes and three major types have been identified (Seligman 1994: 135): expressed, manifest and inventoried. *Expressed interests* are the preferences people report when asked what they like or enjoy. *Manifest interests* are those that are evident from people's lifestyles. *Inventoried interests* are identified by means of a person's pattern of scores on a standardized interest inventory.

A number of goals can be accomplished by interest inventories. The following are mentioned by Seligman (1994: 138-139):

- promote awareness and clarification of interests
- introduce unfamiliar occupations
- increase knowledge of the world of work
- highlight discrepancies between interests and abilities and also between

interests and expressed occupational goals

- translate interests into occupational terms
- organize interests in meaningful and useful ways
- stimulate career thought and exploration
- provide insight into the nature of a person's academic and occupational dissatisfaction
- increase the realism of one's career goals
- reassure people who have already made appropriate tentative career plans
- facilitate conflict resolution and decision making

The results of interest questionnaires, as in the case of personality questionnaires, can be faked in order to make a particular impression. Usually, however, people are more truthful in reporting their interests than in responding to personality questionnaires.

As a general guideline for the *interpretation* of interest questionnaires, the following aspects should be kept in mind:

(1) Interest forms a part of the person's total personality. It involves values, needs, motivation, self-image, etc. Occupational or study counselling thus implies that the entire human personality must be considered.

(2) It should be remembered that interest is not highly correlated with aptitude or ability and that a person who is interested in a specific occupational field or direction of study does not necessarily have the aptitude for it. People with interest patterns not in keeping with their abilities should be helped to develop more realistic occupational aspirations. It is therefore the task of the counsellor to differentiate interest from ability.

(3) Interest can be assumed to be fairly constant beyond the age of about eighteen. The use of interest questionnaires can therefore be fruitful from this age onwards, as long as the possibility of slight variations in the interest

pattern is taken into account.

(4) The close relationship between a person's needs and interests should be kept in mind during interpretation. If needs change, the interest pattern may also change. Interests may, however, also be an indication of a person's needs.

(5) One's value system plays a role in determining one's interests. For example, one may have the interest and ability necessary to succeed in a certain field but on the basis of one's value system nevertheless choose another professional field or direction of study. One's (and especially a pupil's) choice of occupation or field of study will normally correspond to certain value systems which are accepted by one's parents, one's environment and oneself. It is possible, however, to make a choice which conflicts with the value system of those around one. The reason for one's choice may be realistic, i.e. one may have a strong talent and interest in this specific field. The choice may, however, also be unrealistic as a result of identification problems, rebelliousness or immature stubbornness. The counsellor must therefore inform him- or herself about the values of the person involved.

(6) No person can be interested in something about which he or she knows absolutely nothing. Knowledge of occupations can lead to greater or lesser interest. Vocational information thus forms an integral and indispensable part of occupational or study counselling and it is extremely important that both the counsellor and the client have such information at their disposal if they are to do the counselling session justice.

The interest questionnaire can be used to determine provisional occupational options by enabling the general grouping of a person's current interests. A pupil in Standard 10 may be interested in a scientific field. According to his achievements in the relevant school subjects, his interest seems to be realistic. The scientific field offers a wide variety of occupations to choose from. The pupil should begin by making a study of as many of these

occupational choices as possible.

(7) There are three important ways of interpreting interest questionnaires:

- Fields of interest should be grouped into broad interest directions since these provide a better indication of suitable occupations than do separate fields.

- High and low scores must be taken into account during interpretation. Important information can be gained from looking at a person's aversions. A person may, for example, show a strong interest in law, but very little interest in public life.

- Probably no-one will find all his or her interests satisfied by a single career. Some interests must necessarily be practised in the form of hobbies or other activities.

(8) The individual's stated occupational or study preference(s) should always serve as an important point of departure for the interpretation of interest questionnaires. For example, it is rare that a Standard 10 pupil has not considered one option or another, however unrealistic. Such information can be used to gain insight into a pupil's maturity with respect to career choices.

(9) A person's interest profile can also be compared with those of others interested in similar occupations so as to identify similarities. It should, however, be remembered that an individual's profile will not necessarily agree with that of a group of people in a certain occupation. Corresponding trends should thus be sought and not corresponding scores.

(10) When interpreting the results of an interest questionnaire, it must be remembered that the interest patterns of males and females are not necessarily the same.

In conclusion, interest inventories should be interpreted in such a way as to broaden options rather than reinforce stereotyped roles. Interests are relatively stable from age 18 through adulthood and in the hands of a counsellor a good interest questionnaire is therefore a valuable tool for providing career guidance and planning. Although interest does not depend on aptitude or proficiency and is not a consistent predictor of occupational success, the measurement of interest is nevertheless essential in counselling. Only through information regarding the interests of clients can a counsellor help them to identify suitable fields of study or career options that are realistic in the light of their abilities.

7.5 SUMMARY

Almost every day, people have to make decisions about themselves and/or other people. Reliable information on people's knowledge in given areas, their abilities, needs and personality traits, makes the difference between sound and poor decisions, that is, between eventual happiness and frustration.

The fundamental right that each person should have the opportunity to develop his or her abilities and talents fully, to their own benefit as well as that of fellow humans, has an important implication for the educational process. This is that pupils and their parents must often make decisions, which are difficult to change, on the optimalization of the pupils' further education. This holds true for pupils who progress through "normal" education programmes, but is even more relevant for pupils who, for various reasons, may benefit from temporary or permanent special education.

The HSRC has developed a number of instruments which can be used as aids to reduce the uncertainty which necessarily accompanies decisions on optimal education for individuals. The following are available:

Cognitive tests (*inter alia* aptitude and intelligence tests)
A wide range of these tests are available for use by school psychologists. Special student needs and student strengths can be identified. This and other relevant

information can aid decisions with respect to choice of subjects, type of training and careers. Without cognitive tests, human potential cannot adequately be developed.

Personality and related tests and questionnaires

These tests are used by psychologists *inter alia* to diagnose the nature of behavioural disorders and learning difficulties.

Interest questionnaires

These instruments are used for subject and career guidance. Some of the interest questionnaires can be used for self administration and interpretation while others are reserved for professional use.

In conclusion, psychological tests can provide practical solutions to practical problems but they are not infallible. With proper professional control, psychological test results provide information which cannot be obtained more efficiently by other means.

8. PSYCHOLOGICAL TESTING IN SOUTH AFRICA: END OF THE ROAD OR A NEW BEGINNING?

Dark and stormy clouds are gathering which could spell the end of psychological testing in South Africa. There is a perception that "for years, South African psychologists were largely responsible for devising employment tests that were used to screen out blacks from the workplace and from opportunities for development and higher-paying jobs" (Burnette 1994: 8).

Foster, Nicholas and Dawes (1993: 173) mention that the HSRC "was for many years widely held to be no more than a pro-apartheid think-tank". It follows that the psychological instruments produced by this organization were also viewed with suspicion by many. With regard to the National Institute of Personnel Research (NIPR), which was transferred to the HSRC in 1984, Foster *et al.* state that "various commentators have been sharply critical of their research on black personality differences, the use of psychological testing for the exploitation of black labour, and the dominantly instrumentalist perspective of blacks as labour units" (1993: 173). As far as testing is concerned, the same authors are of the opinion that "state-supported psychological testing has left South Africa a legacy of unusable 'race'-based tests" (1993: 173). This is, of course, a over-simplification of a complicated matter. The point is, however, that the authors are conveying a certain impression about tests in South Africa that may be more widespread among test users, including educational authorities, than test developers would like to believe.

A scathing attack on testing in South Africa came from Blade Nzimande (1995), an ANC MP and Chairperson of the Parliamentary Portfolio Committee on Education. In a paper read at a psychometrics conference, he asserted *inter alia* the following:

- In short, testing in South Africa has been fundamentally shaped by apartheid. This therefore begs the question as to whether this basic paradigm has changed significantly.

- What is needed is a kind of internal "truth commission", as part of a scholarly examination of the validity of testing itself. This is even more important given the fact that psychometrics is notorious for its refusal to question the social foundations of its paradigm.

- The implications of a bill of rights for psychological testing are far-reaching. Testing in South Africa developed within the context of national, racial and gender oppression. No matter how much psychologists might have thought they were practising their "science" of testing ..., the fact of the matter is that this was not possible in a society that could be characterised as "unethical".

- The constitution and government are committed to affirmative action as an instrument to redress past historical imbalances. We should therefore pose the question as to whether psychometric development in South Africa is able to grapple with this new reality. This change calls for a complete review of some of the very basic assumptions of psychometrics. ... *The key question facing psychometrics is the analysis of the meaning of affirmative action for testing* (emphasis added).

- The implication of the RDP for testing is that the country is now prioritising human resources development and affirmative action. Testing will have to look at *potential* (emphasis added) and not just actually existing skills in assessing people's capabilities.

- Whilst testing must take into account international developments, it must ultimately be located within the broader social and economic objectives of the society within which it is located.

- Psychology in South Africa is even more American than US psychology itself, and it is this theoretical framework that provides

the paradigmatic basis for testing.

- ... we should question whether testing is needed at all in our conditions. All these years, I have never believed that you can develop culture-free or culture-fair tests, particularly in societies that are characterised by sharp socio-economic divisions and inequalities.

- ... I would like to state that unless testing is able to satisfactorily explore and answer these social questions, I am afraid it is going to be irrelevant and ultimately overtaken by events.

In the above quotations Nzimande makes it quite clear what is in store for psychological testing in South Africa. This, together with the *Green Paper on Employment and Occupational Equity* (1996: 35), which stated that "Employers should avoid psychometric tests unless they can demonstrate that they respect diversity", sounds the death-knell for testing in South Africa as we have known it up till now. Psychometricians who may have thought that this threat was directed only towards occupational tests in an industrial setting are completely mistaken. For two consecutive years now (1995/1996), a committee of the heads of education departments (HEDCOM) failed to grant permission to the HSRC to continue with its research programme in schools which is aimed at the revision of some of the instruments mentioned in paragraph 7 of this document. It may be that the new education authorities are afraid that psychometric testing will serve only to confirm existing inequalities and that some groups may use this outcome as an excuse to justify their claim for "separate" schools - hence HEDCOM's stance on any new developmental work on testing instruments. Unfortunately, the beneficial role tests may play in education, as outlined for example in paragraph 7 of this document, is completely ignored in the whole process.

A second possibility is that, owing to financial constraints, the education departments are not in a position to provide psychological services for all. Therefore, no such luxuries will be provided in future. A third and very likely possibility is that testing, and the instruments developed for this purpose, are

regarded as too "Eurocentric" and that a more indigenous form of assessment must be developed. Whatever the case may be, with the possible exception of testing in the clinical situation by private practitioners, the heyday of psychological testing in schools seems to be over.

The HSRC is trapped in a rather peculiar situation: on the one hand this institution is accused of producing "unusable 'race'-based tests" (Foster *et al.* 1993: 173); on the other hand it is criticized for *not* providing tests that are based on a "particular understanding of the needs of society" (Nzimande 1995: 8), which implies that a test must be culture-specific - therefore culture-fair tests cannot be developed in divergent societies and any attempt to do this will unavoidably result in a biased test! Whatever test developers in this country do, they will inevitably find themselves in a no-win situation.

Fortunately, however, there are still some test *users* who are convinced of the value of tests, especially in clinical usage, and who do not hesitate to caution the new authorities not to reinvent the wheel. One such test user is Shuttleworth-Jordan (1996). In a paper containing sound arguments, she makes an appeal:

> ... against an attitude of nihilism with respect to test usage which occurs because tests have not been designed for application among a particular population, or because appropriate normative data are not yet available. In South Africa this attitude, in its extreme form, promotes a view that all tests in common usage on westernized populations should be abandoned and new culturally relevant and appropriately standardized tests should be designed. In settings dealing with rural and illiterate or semi-literate populations, such a stance has relevance. However, this article cautions against an erroneous exaggeration of cultural effects which fails to take into account the acculturation process. Clinical and research data on urbanized African (Xhosa first language) subjects are used to demonstrate the absence of clinically significant cultural effects on frequently employed, standard test material (1996: 96).

Who will take this advice to heart?

But to return to the question posed in the heading of this paragraph: everything considered, it seems a foregone conclusion that the end of the road for psychological testing (based on the psychometric model) in South Africa is in sight - at least in the case of education departments and the workplace. This disaster for testing can be averted, not by psychometricians as suggested by Nzimande, but only through the actions of the new authorities in education and the labour unions and politicians. It is essential that these and other influential people reconsider their stance on the nature, value and purpose of psychological tests and testing in South Africa. A promising start can be made by abandoning certain cherished rhetoric.

Irrespective of the negative attitudes of some policy makers, there are certain realities the developers of psychometric tests (for educational use) must face. First of all they must realize that the educational scene is very different from what it was two or three years ago. The days of large-scale testing programmes in schools, e.g. the testing of all Std 7 pupils by means of aptitude tests for guidance and counselling with regard to subject choices, are over. Apart from budgetary constraints, there are simply not enough qualified school psychologists to carry out these extensive programmes. In the short to medium term at least, psychological tests will therefore have a limited role to play. This has serious implications for an institution like the HSRC, as there will be no point in embarking on the standardization or re-standardization of psychological tests until the education departments have the capacity to use such tests effectively. In the meantime, however, there is a real danger that the HSRC may lose its capacity to develop tests — unless its researchers can be otherwise meaningfully and productively occupied.

However, all is not lost. A reviewer of an earlier draft of this document (Mr John Brownell) suggests that there is an enormous potential market for curriculum-oriented, teacher-friendly instruments that focus on learning enhancement; these instruments should, however, not be called "psychological" tests. There is a need

for tests where the "psychological" has been de-emphasized so that they can be used by ordinary class teachers, enabling them to assist learners (teachers are the persons most likely to be called upon to administer tests in future). Many of the psychological tests used in education carry restrictions as to their use and interpretation. What is needed, is a refocus on instruments with less restricted use and more direct utility for teachers in terms of influencing what is learned. What is suggested here, is that psychological insights should be used to produce educational tests that focus primarily on assisting teachers to help learners to access the broader curriculum more effectively.

In the same vein another reviewer (Prof. Mervyn Skuy) suggests that the paradigm shift in education — in the form of outcomes-based education — should be mirrored in a paradigm shift in assessment as well. This should include a focus on the individual-environment interaction, and on learning ability, autonomous thinking and functioning, and assessment of potential. He also believes that dynamic assessment should play an increasingly central role in test construction, adaptation, application and interpretation. In this process, consideration should be given to alternative conseptualizations of intelligence with particular reference to the work of Feuerstein (Learning Potential Assessment Device — LPAD), Das (Cognitive Assessment System — CAS) and Kaufman (Kaufman Assessment Battery for Children — K-ABC) — see Par. 4.4 and 4.5. The validity and usefulness of these instruments should be evaluated in the South African context.

In conclusion, although traditional psychometric instruments should play a significant role in education today, their future is anything but secure because of financial constraints, unmanageable counsellor-learner ratios and doubts policy makers have about them. But this does not mean the end of assessment as such. The new kinds of measuring instruments may not be the answer to all assessment problems but at least they are opening up promising new avenues to pursue.

REFERENCES

ADKINS, D.C. 1974. Test construction. Columbus, Ohio: Bell & Howell.

AHMANN, J.S. & GLOCK, M.D. 1959. Evaluating pupil growth. Boston: Allyn & Bacon.

ALLPORT, G.W. 1961. Pattern and growth in personality. New York: Holt, Rinehart & Winston.

ANASTASI, A. 1976. Psychological testing. 4th ed. New York: Macmillan.

ANASTASI, A. 1990. Psychological testing. 6th ed. New York: Macmillan.

AUSUBEL, D.P. 1968. Educational psychology. New York: Holt, Rinehart & Winston.

BALI, S.K., DRENTH, P.J.D., VAN DER FLIER, H. & YOUNG, W.C.E. 1984. Contribution of aptitude tests to the prediction of school performance in Kenya: a longitudinal study. Lisse, The Netherlands: Swets & Zeitlinger.

BINGHAM, W. van D. 1937. Aptitudes and aptitude testing. New York: Harper Bros.

BLOOM, B.S., HASTINGS, J.T. & MADAUS, G.F. 1971. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.

BOUWER, A.C. 1993. Technology for education and training. In R.J. Prinsloo (Ed.), Human Sciences Technology: ways of solving problems in the human domain, pp. 17-53. Pretoria: Human Sciences Research Council.

BROWN, D.C. 1994. Subgroup norming: legitimate testing practice or reverse discrimination? American Psychologist, 49(11), 927-928.

BURNETTE, E. 1994. Psychology's struggle: help heal South Africa. Monitor: American Psychological Association, November.

CARROLL, J.B. 1993. Human cognitive abilities: a survey of factor-analytic studies. Cambridge: Cambridge University Press.

CATTELL, R.B. 1965. The scientific analysis of personality. Harmondsworth: Penguin.

CATTELL, R.B. 1983. The role of psychological testing in educational performance: the validity and use of ability predictions. The Mankind Quarterly, XXIII (Nos. 3 & 4), 227-277.

CECI, S.J. & WILLIAMS, W.M. 1997. Schooling, intelligence, and income. American Psychologist, 52(10), 1051-1058.

CLEARY, T.A. 1968. Test bias: prediction of grades of Negro and White students in integrated colleges. Journal of Educational Measurement, 5(2), 115-124.

CRONBACH, L.J. 1990. Essentials of psychological testing. 5th ed. New York: Harper Collins.

DANIEL, M.H. 1997. Intelligence testing: status and trends. American Psychologist, 52(10), 1038-1045.

DAS, J.P., NAGLIERI, J.A. & KIRBY, J.R. 1994. Assessment of cognitive processes: The PASS theory of intelligence. Needham Heights, MA: Allyn & Bacon.

DAVIDSON, G. 1995. Cognitive assessment of indigenous Australians: towards a multiaxial model. Australian Psychologist, 30(1), 30-34.

DAVIDSON, G. 1996. Fairness in a multicultural society! Reply to Dyck (1996). Australian Psychologist, 31(1), 70-72.

DYCK, M.J. 1996. Cognitive assessment in a multicultural society: comment on Davidson (1995). Australian Psychologist, 31(1), 66-69.

EDIGER, M. 1994. Measurement and evaluation. Studies in Educational Evaluation, 20, 169-174.

FERGUSON, G.A. 1954. On learning and human ability. Canadian Journal of Psychology, 8(2), 95-112.

FEUERSTEIN, R. 1980. Instrumental enrichment: an intervention program for cognitive modifiability. Glenview, ILL.: Scott, Foresman.

FEUERSTEIN, R., RAND, Y. & HOFFMAN, M. 1979. The dynamic assessment of retarded performers: Learning potential assessment device, theory, instruments, and techniques. Baltimore: University Park Press.

FOSTER, D., NICHOLAS, L. & DAWES, A. 1993. A reply to Raubenheimer. The Psychologist, April.

FOUCHÉ, F.A. & VERWEY, F.A. 1978. Manual for the Senior Aptitude Tests 1978 Edition (SAT 78). Pretoria: Human Sciences Research Council.

GARDNER, H. & HATCH, T. 1989. Multiple intelligences go to school. Educational Researcher, 18(5), 4-9.

GEISINGER, K.F. 1994. Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. Psychological Assessment, 6(4), 304-312.

GEKOSKI, N. 1964. Psychological testing. Springfield, ILL.: Charles C. Thomas.

GHUMAN, P.A.S. 1980. A study of the concept of equivalence and divergent thinking among four subcultural groups of Punjabi children. International Review of Applied Psychology, 29, 89-103.

GOODWIN, W.L. & DRISCOLL, L.A. 1980. Handbook for measurement and evaluation in early childhood education. San Francisco: Jassey-Bass.

GOTTFREDSON, L.S. 1994. The science and politics of race-norming. American Psychologist, 49(11), 955-963.

GREEN, J.A. 1970. Introduction to measurement and evaluation. New York: Dodd, Mead & Co.

GREEN PAPER ON EMPLOYMENT AND OCCUPATIONAL EQUITY, 1996. Government Gazette, 1 July.

GUTHRIE, G. 1963. Structure of abilities in a non-western culture. Journal of Educational Psychology, 54, 94-103.

HAMERS, J.H.M. & RESING, W.C.M. 1993. Learning potential assessment: introduction. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijssenaars (Eds.), Learning potential assessment (pp.23-41). Lisse: Swets & Zeitlinger.

HAY, J.F. & PIETERS, H.C. 1994. The interpretation of large differences in a psychiatric population between verbal IQ (VIQ) and non-verbal IQ (NVIQ) scores when using the Senior South African Individual Scale (SSAIS). In R. van Eeden, M. Robinson & A.B. Posthuma (Eds.), Studies on South African Individual Intelligence Scales, pp. 123-138. Pretoria: Human Sciences Research Council.

HELMS, J.E. 1992. Why is there no study of cultural equivalence in standardized cognitive ability testing? American Psychologist, 47(9), 1083-1101.

HJELLE, L.A. & ZIEGLER, D.J. 1976. Personality theories: basic assumptions, research and applications. New York: McGraw-Hill.

HOY, C. & GREGG, N. 1994. Assessment: the special educator's role. Pacific Grove, Calif.: Brooks/Cole.

JENSEN, A.R. 1980. Bias in mental testing. New York: Free Press.

KAUFMAN, A.S. & KAUFMAN, N.L. 1983. Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service.

KEHOE, J.F. & TENOPYR, M.L. 1994. Adjustment in assessment scores and their usage: a taxonomy and evaluation of methods. Psychological Assessment, 6(4), 291-303.

KLINE, P. 1993. The handbook of psychological testing. London: Routledge.

LABORATORY OF COMPARATIVE HUMAN COGNITION. 1979. What's cultural about cross-cultural cognitive psychology? Annual Review of Psychology, 30, 145-172.

LESSER, G.S., FIFER, G. & CLARK, D.H. 1965. Mental abilities of children from different social-class and cultural groups. Monographs of the Society for Research in Child Development, 30(4), 1-115.

LI, R. 1996. A theory of conceptual intelligence: thinking, learning, creativity and giftedness. Westport, Conn.: Praeger.

LIDZ, C.S. 1997. Dynamic assessment approaches. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), Contemporary intellectual assessment (pp.281-296). New York: The Guilford Press.

LURIA, A.R. 1973. The working brain. New York: Basic Books.

MACKENZIE, V. 1981. Testing minority group children. Australian Psychologist, 16(2), 234.

MEHRENS, W.A. & LEHMAN, I.J. 1973. Measurement and evaluation in education and psychology. 2nd ed. New York: Holt, Rinehart & Winston.

MESSICK, S. 1982. The values of ability testing: implications of multiple perspectives about criteria and standards. Princeton: Educational Testing Service.

MILLER-JONES, D. 1989. Culture and testing. American Psychologist, 44(2), 360-366.

MOORE, E.G.J. 1987. Ethnic social milieu and black children's intelligence test achievement. Journal of Negro Education, 56(1), 44-52.

NAGLIERI, J.A. 1997. Planning, attention, simultaneous, and successive theory and the Cognitive Assessment System: a new theory-based measure of intelligence. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), Contemporary intellectual assessment (pp.247-267). New York: The Guilford Press.

NZIMANDE, B. 1995. To test or not to test? Paper delivered at the Psychometrics Conference, held at Pretoria 5 - 6 June.

OAKLAND, T. 1995. 44-Country survey shows international test use patterns. Psychology International, 6(1), 7.

OLSON, D.R. 1986. Intelligence and literacy: the relationship between intelligence and the technologies of representation and communication. In R.J. Sternberg & R.K. Wagner (Eds.), Practical intelligence: Nature and origins of competence in the everyday world, pp. 338-360. Cambridge: Cambridge University Press.

OWEN, K. 1991. Test bias: the validity of the Junior Aptitude Tests (JAT) for various population groups in South Africa regarding constructs measured. South African Journal of Psychology, 21, 112-118.

OWSTON, R.D. 1984. Detecting bias in standardized tests: a suggested procedure for counsellors. School Guidance Worker, 39(4).

PETERSEN, N.S. 1980. Bias in the selection rule - bias in the test. In L.J.T. van der Kamp, W.F. Langerak & D.N.M. de Gruijter (Eds.), Psychometrics for educational debates. Chichester: John Wiley.

PLUG, C., MEYER, W.F., LOUW, D.A. & GOUWS, L.A. 1986. Psigologie-woordeboek. 2de uitgawe. Johannesburg: McGraw-Hill.

RESCHLY, D.J. 1997. Diagnostic and treatment utility of intelligence tests. In D.P. Flanagan, J.L. Genshaft & P.L. Harrison (Eds.), Contemporary intellectual assessment (pp.437-456). New York: The Guilford Press.

RESNICK, L.B. & RESNICK, D.P. 1982. Testing in America: the current challenge. International Review of Applied Psychology, 31(1), 76-90.

REYNOLDS, C.R. 1982. The problem of bias in psychological assessment. In C.R. Reynolds & T.B. Gutkin (Eds.), The handbook of school psychology, pp. 178-208. New York: Wiley.

REYNOLDS, C.R. & BROWN, R.J. 1984. Bias in mental testing: an introduction to the issues. In C.R. Reynolds & R.J. Brown (Eds.), Perspectives on bias in mental testing. New York: Plenum Press.

RUSSEL, R.W. & CRONBACH, L.J. 1958. Report of testimony at a congressional hearing. American Psychologist, 13, 217-223.

SACKETT, P.R. & WILK, S.L. 1994. Within-group norming and other forms of score adjustment in preemployment testing. American Psychologist, 49(11), 929-954.

SCARR, S. 1981. Race, social class and individual differences in IQ. Hillsdale: Lawrence Erlbaum.

SCHMIDT, F.L. & HUNTER, J.E. 1981. Employment testing: old theories and new research findings. American Psychologist, 36(10), 1128-1137.

SELIGMAN, L. 1994. Developmental career counseling and assessment. 2nd ed. Thousand Oaks, Calif.: Sage.

SEMEONOFF, B. 1966. Personality assessment. Harmondsworth: Penguin.

SHUTTLEWORTH-JORDAN, A.B. 1996. On not reinventing the wheel: a clinical perspective on culturally relevant test usage in South Africa. South African Journal of Psychology, 26(2), 96-102.

SINGHAM, M. 1995. Race and intelligence: what are the issues? Phi Delta Kappan, December, 271-278.

STAGNER, R. 1974. Psychology of personality. 4th ed. New York: McGraw-Hill.

STANLEY, J.C. & HOPKINS, K.D. 1972. Educational and psychological measurement and evaluation. 5th ed. Englewood Cliffs: Prentice-Hall.

STERNBERG, R.J. 1997a. The concept of intelligence and its role in lifelong learning and success. American Psychologist, 52(10), 1030-1037.

STERNBERG, R.J. 1997b. Intelligence and lifelong learning: what's new and how can we use it? American Psychologist, 52(10), 1134-1139.

SUZUKI, L.A. & VALENCIA, R.R. 1997. Race-ethnicity and measured intelligence: educational implications. American Psychologist, 52(10), 1103-1114.

THORNDIKE, R.L. & HAGEN, E. 1969. Measurement and evaluation in psychology and education. 3rd ed. New York: Wiley.

THORNDIKE, R.M., CUNNINGHAM, G.K., THORNDIKE, R.L. & HAGEN, E.P. 1991. Measurement and evaluation in psychology and education. 5th ed. New York: Macmillan.

TUCKMAN, B.W. 1975. Measuring educational outcomes: fundamentals of testing. New York: Harcourt Brace Jovanovich.

TYLER, R.W. 1966. General statement on evaluation. In C.I. Chase & H.G. Ludlow (Eds.), Readings in educational and psychological measurement. Boston: Houghton Mifflin.

VALENCIA, R.R. & LOPEZ, R. 1992. Assessment of racial and ethnic minority students: Problems and prospects. In M. Zeidner & R. Most (Eds.), Psychological testing: an inside view, pp. 309-439. Palo Alto, Calif.: Consulting Psychologists Press.

VAN DER WESTHUIZEN, J.G.L. 1979. Manual for the use of psychological and scholastic tests as aids in school guidance. Pretoria: Human Sciences Research Council.

VAN EEDEN, R. 1993. The validity of the Senior South African Individual Scale-Revised (SSAIS-R) for children whose mother tongue is an African language: private schools. Pretoria: Human Sciences Research Council.

VYGOTSKY, L.S. 1978. Mind in society: The development of higher-order psychological processes. Cambridge, MA: Harvard University Press.

WHITE, S.W. 1988. Opportunity and intelligence. National Forum: Phi Kappa Phi Journal, 68, 2-3.

WOOD, R. 1986. Aptitude testing is not an engine for equalising educational opportunity. British Journal of Educational Studies, XXXIV (1), 26-37.

ZAPPARDINO, P. 1995. Science, Intelligence, and Educational Policy: the mismeasure of Frankenstein (with apologies to Mary Shelley and Stephen Jay Gould). Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, April 18-22. Document ED 384621.

ZEIDNER, M. & MOST, R. (Eds.), 1992. Psychological testing: an inside view. Palo Alto, Calif.: Consulting Psychologists Press.

ZINDI, F. 1995. Intelligent or not? The African's dilemma. SAPEM, August.

01404G

HSRC
RGN
Group Education

BEST COPY AVAILABLE